

## A blast from the past

**NASA is right to invest heavily in nuclear propulsion to speed the progress of future missions to the outer planets. But it may pay to complete our initial exploration of the Solar System the old-fashioned way.**

**I**t seems unfair to call the Cassini space probe — celebrated in this week's issue (see pages 965–966 and 985–1005) — a relic of the past when it is still two years away from the encounter with Saturn that is its *raison d'être*.

Cassini is the most capable deep-space explorer ever launched, with optics and science instruments that are far superior to those on board its predecessors, such as the Galileo mission to Jupiter and its moons. The show from Saturn and its mysterious, cloud-covered moon Titan should be spectacular. And, unlike Galileo, Cassini isn't crippled by a broken antenna, so we should receive every byte of data that its design engineers planned for.

Still, the \$3.5-billion, six-tonne Cassini is like a 1959 Cadillac: heavy, expensive and a thing of beauty — but outdated. We would do things differently today than we did in 1989, when the project left the drawing board.

Just how differently was hinted at earlier this month, when NASA announced its plan to invest substantially in space nuclear propulsion for the first time in years. This is great news for planetary science. New propulsion systems could dramatically cut travel times to the outer planets and enable an exciting new class of missions, such as placing an orbiter around Europa, the possibly life-bearing moon of Jupiter.

Caught between the old and new ways of doing business is NASA's

proposed mission to Pluto, which, like Titan, is one of the Solar System's last bits of true terra incognita. The agency has been struggling for nearly a decade with the problem of how best to explore this distant world. Last year, at the US Congress's insistence, it chose a concept called New Horizons, which uses conventional rocket propulsion, for a launch in 2006. Now that plan is on hold, while NASA revamps its outer-planets programme to incorporate possible new propulsion technologies and waits for scientific advice on which planetary targets are the most attractive.

A mission to Pluto and the Kuiper Belt — the vast population of chunks of rock and ice that lies beyond the orbit of Neptune — is likely to be high on that list. But tempting though it may be to speed up the journey with nuclear rockets, it may ultimately make more sense to stay with the proposal already on the table. A revolutionary new propulsion system will take several years to develop and test in space. And while a nuclear rocket would be fast, reaching Pluto in as little as five years, it would not be cheap.

Meanwhile, the New Horizons team has cut its proposed travel time to under ten years, and would spend less than \$500 million — well below the \$650-million cap for the new breed of outer planetary missions. So while NASA should vigorously pursue the nuclear option, the best way to reach the last uncharted planet, at least in the short term, is probably still the old-fashioned way. ■

## Burdened by expectation

**Japan is asking too much of its latest flagship initiative in the biological sciences.**

**B**y any reckoning, the Center for Developmental Biology (CDB) that is soon to open its doors in Kobe, Japan (see pages 952–953), is a bold initiative. Its facilities are second to none, and its leaders have pledged to give talented young scientists free rein to develop their own research interests. Given that rigid academic hierarchies have repeatedly been blamed for Japan's perceived underachievement in basic science, expectations are high.

The danger, however, is that they may be too high, and that the CDB will be pressured into trying to advance on more fronts than is reasonable.

Increasingly, for instance, Japan has made overtures to recruit scientists from overseas, especially from the West, with limited success. The CDB aims to have a large number of foreigners on its roster, but this may not be a wise place to focus its efforts. There are many obstacles beyond the institute's control, such as language, a dearth of reasonably priced international schools, and fears that women will not be treated as equals.

The CDB also has to deal with the short-term vision of the current government, which wants the institute to become the hub of a cluster of bioscience companies. This is a laudable aim — but again, the CDB may have its work cut out achieving this goal.

There are many who feel that, lacking the freewheeling entrepre-

neurial spirit that pervades the labs of California, Japan will never boast a vibrant biotech industry.

The CDB's director, Masatoshi Takeichi, is already worried that short-term pressures will encourage the centre's scientists to shoot for some headline-grabbing discovery, rather than building the solid foundations of a project that will truly advance their field.

Rather than worrying about the initial roster of foreign recruits, or fretting about commercial spin-offs, the CDB's paymasters would be wise to back Takeichi in his goal of making the CDB into a world leader in advancing the fundamental scientific knowledge needed to underpin the emerging field of regenerative medicine. A reasonable target would be to aim to produce a defined number of landmark papers in areas such as stem-cell biology, transgenic technology, wound healing and cellular imaging.

If the CDB can establish a reputation for excellence in these areas, and its group leaders can resist the temptation to become as autocratic with their own charges as their professors may in the past have been with them, there may no longer be a need to worry about meeting a quota of foreigners. And if young researchers from around the world are then inspired to beat a path to the CDB's door, they may also bring with them some of the entrepreneurial skills that the Japanese government is so keen to promote. ■





**In the spotlight**  
Standards proposed  
for microarray  
research  
p946



**Reef encounter**  
Marine biologists  
gather to save coral  
from bleaching  
p947



**Corn trouble**  
Row escalates over  
use of genetically  
modified crops  
p948



**Linus online**  
Pauling's notebooks  
set to make their  
Internet debut  
p950

# Nuclear-weapons design plan raises fresh proliferation fears



BEN MARGOT/AP

Livermore's ASCI White computer can be used to simulate nuclear weapons — old and new.

**Geoff Brumfiel, Washington**

The Bush administration has announced plans for the United States to resume design work on new nuclear weapons for the first time in almost a decade.

The move drew immediate fire from former nuclear-weapons designers, scientific organizations and environmental groups, who say it could undermine efforts to limit the spread of nuclear weapons.

Herbert York, a former weapons designer and arms-control advocate, says the plan has "the wrong kind of focus". York says the nuclear-weapons programme should instead be working to minimize the chances that the weapons will ever be used.

The plan was announced by John Gordon, director of the National Nuclear Security Agency (NNSA), the branch of the Department of Energy that runs the US weapons programme. "The vision is for small, focused teams to assess military requirements, investigate options and ensure that our Department of Defense partners understand what is and is not possible," Gordon told a hearing of the Senate Committee on Armed Services on 14 February.

Gordon said that, beyond assessing the needs of the armed forces, the programme will start design work on new kinds of nuclear

warheads. John Harvey, director of planning at the NNSA, says: "What we're asking these teams to do is to think about what kind of things will be possible in the future."

The emphasis on weapons development is a sharp departure from the approach taken by the Clinton administration, which, from 1992 to 2000, told the weapons laboratories to concentrate on maintaining existing weapons, under the Science-Based Stockpile Stewardship programme (see *Nature* **387**, 541; 1997). The design work during that period was largely restricted to modifying the weapons systems that carried the warhead, according to government statements.

The new plan calls for the establishment of small design teams at three US nuclear weapons laboratories — Los Alamos and Sandia in New Mexico, and Lawrence Livermore in California. Their initial focus will probably be on small, earth-penetrating warheads, in which the defence department has expressed an interest. The labs have already developed such a weapon, based on existing nuclear warheads.

The NNSA claims that the design teams are being formed largely to help train and recruit scientists and engineers. "A key element of this is to take the scientists who have developed, designed and tested new

## Foreign researchers turn their backs on Germany

**Quirin Schiermeier and Regina Wegner, Munich**  
Germany continues to be an unpopular destination for young researchers from abroad, according to new figures from the European Commission (EC).

The country's lack of attractiveness to researchers has been revealed in details of the uptake of the EC's Marie Curie Fellowships, which enable European post-graduates and postdoctoral fellows to study in another country within the continent.

Only 10% of the 2,080 fellowships awarded between 1999 and 2001 were used to study at a German university, research institute or industry research programme.

The Netherlands, a country with a much smaller science base than Germany, attracts almost the same number. Britain remains the preferred country for young European scientists, attracting over one-third of new fellows.

France is a relatively popular destination, drawing 17% of new fellows. But, together with Spain, the country also has the largest number of researchers departing through the scheme. Scientists from the two nations accounted for 40–50% of all new fellowships between 1999 and 2001. Both countries have a relatively small number of new postdoctoral positions available each year.

German researchers contacted by *Nature* say that language difficulties and the notorious bureaucracy of university administrations are to blame.

The situation is particularly difficult in eastern Germany, says Jörg Oehlmann, a toxicologist at the University of Frankfurt, who has supervised Marie Curie fellows in the past. "Many people are deterred by the fact that English is only poorly spoken and understood in eastern Germany, which can make daily life a rather adventurous experience for foreigners," he says.

► [www.cordis.lu/improving/fellowships/home.htm](http://www.cordis.lu/improving/fellowships/home.htm)

warheads, and have them transfer their skills to a new generation," says Harvey. Maintaining these skills will allow the United States to respond to future threats, he adds.

But Chris Paine, an analyst at the Natural Resources Defense Council, an environmental group that keeps a close track of the US weapons programme, says that such skills are unnecessary. "This is no longer a growth industry," he says. Paine predicts that the design effort will serve to undermine non-proliferation, and encourage other countries to develop nuclear weapons of their own.

Richard Garwin, a physicist and former head of research at IBM who has advised successive US governments on nuclear-weapons policy, suggests that the design activity may ultimately lead to a resumption of nuclear testing, which the United States abandoned in 1992.

But Harvey maintains that most of the activity would be restricted to computer simulation and testing of weapon components. "Our intention would be to carry out new development consistent with the president's moratorium on testing," he says. But he concedes that the possibility of future tests remains "an open question." ■

## Power vacuum expands as CDC director resigns

**Meredith Wadman, Washington**

Four of the main health administration jobs in the US government are now vacant, following Jeffrey Koplan's resignation as director of the Centers for Disease Control and Prevention (CDC) on 21 February.

As well as the CDC, the National Institutes of Health (NIH) and the Food and Drug Administration are without permanent directors. The figurehead position of surgeon general — who offers health advice to the US public — fell vacant a few weeks ago with the departure of David Satcher.

The vacancies are a source of mounting concern to health advocates, who say that the situation will take its toll on biomedical research and public health. "It's just a glaring gap in scientific leadership at a critical time," says Tony Mazzaschi, associate vice-president for research at the Association of American Medical Colleges.

But Bill Hall, a spokesman for health secretary Tommy Thompson, says: "We have

very, very competent people who are acting in those positions right now and who are doing a fantastic job."

Nevertheless, the NIH has been without a permanent director for more than two years. And in the past week Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases (NIAID), has fallen out of the running for the NIH's top job. White House officials say this was because he wanted to maintain his job at the NIAID while serving as NIH director — but others contend that conservatives spiked his nomination because of his past support for fetal-tissue research.

Elias Zerhouni, a radiologist and senior administrator at Johns Hopkins University School of Medicine in Baltimore, Maryland, is now being floated as a contender for the post. But he inspired a new centre at Johns Hopkins that conducts embryonic stem-cell research, and looks as vulnerable as anyone to the political machinations that are holding up health-related appointments. ■

## Minimum standards set out for gene-expression data

**Jonathan Knight, Boston**

A grassroots collection of biologists plans to persuade scientific journals to enforce minimum standards for the publication of all experiments involving microarrays.

The arrays, which monitor the expression of thousands of genes simultaneously, have proved to be helpful in the discovery of genes that influence such diverse processes as cancer and embryonic development. But whereas DNA sequences are always in the same format, microarray

experiments are difficult to compare with one another, because of the multitude of variables that affect their outputs (see *Nature* 410, 851, 860–861; 2001).

In an effort to remedy this, the Microarray Gene Expression Database (MGED) Group first devised a common set of standards — called minimum information about a microarray experiment (MIAME) — in 1999.

But although many papers now refer to the MIAME standards, compliance is uneven, says Catherine Ball, curator of the Stanford Microarray Database. "At the moment, it's terrible," she says. "I often find I can't identify genes or that people have used home-grown software they don't describe."

"Generally, we want the same thing you have with published DNA sequences," says Alvis Brazma, head of microarray informatics at the European Bioinformatics Institute in Cambridge, UK. This means that all primary research data should be submitted to a repository, he says. The standards would also require the publication of enough experimental detail to allow other scientists to compare different data sets.

In early March, the MGED steering committee, of which Ball and Brazma are members, plans to release a check-list based on MIAME for use by authors, editors and

referees. The list will be designed to make it easy for journals to request, and ultimately require, MIAME compliance.

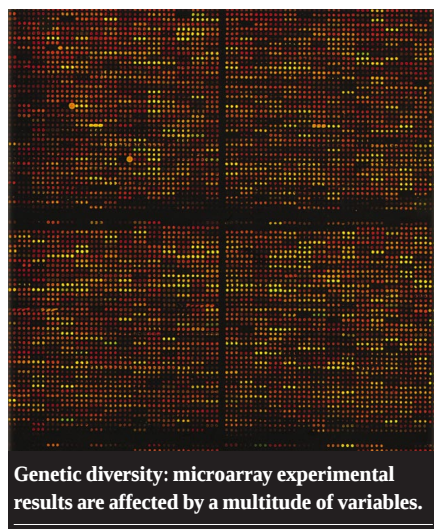
The final check-list is likely to ask for details of sample preparation and data processing, as well as numerical read-outs of the intensity of the array's red and green spots, which indicate the level of gene expression.

Ball says that the check-list won't require authors to reveal every gene on an array, as this could allow others to scoop experiments in progress. "No one would argue that you should publish data prematurely," she says. "You can strip out the gene identity of things you are not publishing."

Journal editors are enthusiastic about the proposed check-list, but say that they will consult the microarray researchers before requiring authors to comply with it. "We'd consider insisting if the community showed broad consensus," says *Nature's* editor, Philip Campbell.

Laurie Goodman, executive editor of *Genome Research*, published by Cold Spring Harbor Laboratory, says that she would be cautious about requirements that could rule out important papers, but would consider adopting minimum standards if she received a strong indication that microarray users support them. ■

♦ [www.mged.org](http://www.mged.org)



Genetic diversity: microarray experimental results are affected by a multitude of variables.



# Reef under threat from 'bleaching' outbreak

Carina Dennis, Sydney

Coral experts from around the world are meeting on the Great Barrier Reef to hammer out a research strategy in response to coral 'bleaching'.

This year's episode "may turn out to be the worst bleaching event on the Great Barrier Reef in recent history", says Ove Hoegh-Guldberg, director of the Centre for Marine Studies at the University of Queensland, who is chairing a three-week workshop on the problem at Heron Island Research Station, 80 kilometres off the Australian mainland.

Mass coral bleaching — estimated to have destroyed about one-sixth of the world's coral colonies during the last major occurrence in 1998 — is attributed by researchers to mild increases in ocean temperature (see below).

Coral death so far this year has been confined to reefs close to the northeast Australian coast. But according to Paul Marshall, a project manager at the Great Barrier Reef Marine Park Authority, "we are still a long way away from the end of summer and if conditions don't improve substantially, we will start to see severe bleaching and mortality over extensive areas of the reef".

Mass bleaching typically occurs when oceans are warmed by the short-term El Niño climate effect. But the 2002 bleaching is out of phase with El Niño, raising concerns that bleaching events are growing in frequency and intensity in response to climate change. "The only thing we can point at is global warming," says William Skirving of the Australian Institute of Marine Science (AIMS).

Skirving's research team has developed computer models to help to predict the susceptibility of reefs to bleaching. Using information on sea-surface temperature, water depth and mixing arising from currents and tides, they can produce maps of bleaching risk, showing which parts of the reef are most vulnerable. The AIMS model complements methods developed by Al Strong of the US National Oceanic and Atmospheric Administration (NOAA), which use satellite monitoring of sea-surface temperature.

Although remote-sensing techniques are aiding the monitoring of bleaching events, biologists have limited knowledge of the physiological and ecological consequences of bleaching. "We still have a very small understanding of what proportion of bleached coral dies and what bleaching means for the ecosystem," says Marshall.

"We know the times and temperatures required to induce coral bleaching, but the question now is how much more does it take for corals to die," says Ray Berkemans of AIMS, who is experimentally stressing corals in the laboratory to determine thresholds for bleaching. More knowledge of changes in the cellular metabolism and composition of coral during temperature shifts — for example, altered expression of 'heat-shock' proteins — could provide powerful physiological indicators for monitoring bleaching, he adds.

An important question is whether coral can adapt to increased temperatures. There has been much debate over whether coral bleaches as part of an adaptive strategy, including an exchange in *Nature* earlier this month (see *Nature* 415, 601–602; 2002). The "adaptive-bleaching hypothesis" suggests that temperature-stressed corals bleach in order to adopt more heat-tolerant varieties of algae — a scenario dismissed as over-optimistic by Hoegh-Guldberg and others.

Berkemans says he has found little evidence of thermal adaptation after years of experiments in which he transplanted corals from cool areas to warm areas. "Even after 12 months in a warmer environment, the corals still show responses that are equivalent to that in their native cooler climate," he says.

If coral adaptation cannot keep pace with increasing sea temperatures, the survival of the world's reefs is in jeopardy. "We need to coordinate research across international boundaries," says Hoegh-Guldberg. That is the challenge for the 45 researchers gathered at the Heron Island meeting, which was convened by the Intergovernmental Oceanographic Commission of UNESCO and the World Bank. The group plans to establish a global network of researchers within six months to find answers to the mysteries that still surround coral bleaching. ■



**White plight:** researchers fear that Australia's corals may be devastated by bleaching this year.



## How corals lighten up

Coral bleaching occurs when stressed coral expels the symbiotic algae embedded in it. The algae give coral its brown colour, so that when the algae are lost, the coral's calcium carbonate skeleton is exposed, rendering the reef white.

The dinoflagellate algae, known as zooxanthellae, provide the coral with energy and nutrients from their photosynthetic processes; in return, corals supply algae with essential compounds, such as ammonia and phosphate, from their waste metabolism. If the algae do not recolonize the coral tissues rapidly enough, the coral

starves and dies.

Incidences of local bleaching can be spurred on by a range of environmental and physiological stresses, such as heat, solar radiation, pollution, reduced salinity and changes in oxygenation. But mass bleaching is primarily caused by elevated temperatures. Corals live close to their upper thermal limit and when ocean temperatures rise 1 °C or higher above long-term monthly averages for the warmest months, bleaching begins. Periods of still winds, clear skies and little current during the summer months on the reef are peak times for bleaching.

It is not fully understood why algae are expelled from stressed coral, but it is thought that the algae become toxic to the coral at higher temperatures. It has been proposed that elevated temperatures upset the photosynthetic metabolism in the algae, leading to the production of toxic free radicals. Researchers still debate whether it is primarily the light or dark reactions of photosynthesis that are affected.

Although coral may survive a bleaching event and recover their algae, their reproductive capacity is nonetheless impaired, resulting in long-term damage to the coral system. **C.D.**

# Alleged flaws in gene-transfer paper spark row over

## Declan Butler

A political row has erupted over a scientific paper by authors who claim to have found transgenic DNA from genetically modified (GM) maize in local varieties of the crop in Mexico.

Calls from environmental groups to halt the planting of transgenic crops in Mexico and elsewhere followed hot on the heels of the paper's publication in *Nature* last November (414, 541–543; 2001). But some researchers have since raised questions over the study's validity.

The paper was written by Ignacio Chapela and David Quist of the University of California, Berkeley. In it they report that a promoter sequence of DNA that originated in the cauliflower mosaic virus has shown up in creole maize varieties in two remote mountain areas of Mexico. The viral DNA is used in a variety of GM maize to enhance the activity of the introduced genes.

Their results led the authors to ask whether such gene flow would threaten native species of maize in Mexico, the centre of origin for the crop.

But an editorial in this month's *Transgenic Research* (11, 3–5; 2002) says that "the data presented in the published article are mere artifacts resulting from poor experimental design and practices". The article was written on behalf of the journal's

editorial board by the editor, Paul Christou, who is director of the molecular biotechnology unit at the John Innes Centre in Norwich, UK. It concludes that "no credible scientific evidence is presented in the paper to support claims made by the authors".

Philip Campbell, editor of *Nature*,

declines to discuss the matter in detail. "We treat all submissions as confidential and so are not willing to comment on specific cases," he says. "Our policy in general is to consider criticisms received after publication as promptly as possible."

But on 19 February, 144 non-governmental organizations, led by the



Concerns about gene flow run strong in Mexico, the centre of origin for natural maize.

## Academy proposes tighter crop monitoring

### Virginia Gewin, Washington

The US Department of Agriculture should engage in much broader monitoring of the environmental consequences of new crop types, whether they are genetically modified or not, according to a study by the National Academy of Sciences.

The study criticizes the Animal and Plant Health Inspection Service (APHIS), the arm of the agriculture department responsible for crop monitoring, for lacking transparency and scientific expertise. It says the service needs to recruit more ecologists and set up an external scientific advisory panel to enhance the rigour of proposed regulatory changes.

Supporters of transgenic crops argue that special rules are unnecessary because such crops carry no special risks. But the Academy study — written by a panel chaired by Fred Gould, an entomologist at North Carolina State University — introduces a new twist to this argument, by declaring that conventional crops may also pose environmental risks.

The study's authors say that conventional breeding techniques, such as mutagenesis and embryo rescue, also involve genetic

manipulation and so should be monitored. "Conventional plant breeding is not what many people think it is," says Alan McHughen, a geneticist at the University of California, Riverside, and a member of the study panel.

The study suggests that the potential environmental effects of conventionally bred crops should be re-evaluated. It also calls for environmental monitoring of crops after commercialization, to ensure that earlier risk assessments have been accurate.

Jane Rissler, a director of the Union of Concerned Scientists, welcomes the study, saying that it identifies "areas where the agriculture department needs to significantly improve if it wants to regulate the products of the future". She contends that the current system makes it too easy for crops to enter commercial use.

But Michael Phillips, executive director for food and agriculture at the Biotechnology Industry Organization, says that thousands of products have been introduced so far "without any significant harm either to food safety or to the environment". He claims that the agriculture department's existing

regulatory system is viewed as "the gold standard" around the world.

The study was originally requested by Dan Glickman, agriculture secretary in the Clinton administration, and the Bush administration is not considered likely to act on its recommendations. APHIS said in a statement that it had already addressed many of the issues raised, taking steps, for example, to set up a scientific advisory panel. ■





## genetically modified maize

Canada-based Action Group on Erosion, Technology and Concentration (ETC), issued a statement alleging that “pro-industry academics are engaging in a highly unethical mud-slinging campaign against the Berkeley researchers”.

The statement calls on the UN Food and Agriculture Organization and the Consultative Group on International Agricultural Research (CGIAR), which runs 16 agricultural research laboratories around the world, to “propose an immediate moratorium on the shipment of GM seed or grain in countries or regions that form part of the centre of origin or centre of genetic diversity for the species”.

Alex Avery, director of research at the Virginia-based Center for Global Food Issues, which supports agricultural biotechnology, dismisses the ETC’s statement as “some last-minute damage control before the Quist and Chapela study is thoroughly refuted”. He claims that the ETC “doesn’t care about the scientific debate — it is just trying to sway reporters into bolstering the credentials of Chapela and Quist”.

Both sides of the argument are hoping to influence imminent decisions about the regulation of transgenic crops — in particular, the fate of the existing European

Union moratorium on their commercial use. The United States, which views the ban as protectionism, would like to see its removal discussed at the next meeting of European leaders in Barcelona on 15–16 March. But according to one official at the European Commission, a decision to lift the moratorium is unlikely until the end of the year at the earliest.

Some experts say that the debate on the Mexico findings is, in any case, somewhat beside the point. Because maize is wind-pollinated and varieties cross readily, almost everyone agrees that genes from GM maize will cross to local varieties if they are grown close together.

What really matters is the ecological impact of such gene flow. Local maize varieties are not very stable, and farmers have long crossbred them with other varieties. “Gene flow is a constant,” says Tim Reeves, director of the CGIAR International Maize and Wheat Improvement Center (CIMMYT) in Mexico. “The real question is whether it makes any difference if one of the genes that flows in is a transgene.”

Scientists are divided on that question. Some argue that the transgenes will reduce genetic diversity, whereas others contend that they could either have a neutral effect or actually enhance diversity. ■

## Poor nations seek new biodiversity deal

Virginia Gewin, Washington

Twelve countries in the developing world have issued a declaration requesting an international agreement to ensure the fair and equitable sharing of benefits derived from the world’s biodiversity, such as new drugs and agricultural products.

Between them, the countries — which include China, Mexico, India and Brazil — claim to house nearly three-quarters of the Earth’s biodiversity.

The declaration — also signed by Indonesia, Costa Rica, Colombia, Kenya, Ecuador, Peru, Venezuela and South Africa — was issued on 18 February at a meeting of environment ministers in Cancun, Mexico.

The declaration is seen as an attempt by the nations to stake out a unified position ahead of two international meetings — a conference of the parties to the Convention on Biological Diversity (CBD) in The Hague in April, and the United Nations’ World Summit on Sustainable Development in Johannesburg, South Africa, in August.

The CBD was set up in 1992 to create a structure for agreement between governments and corporations for the equitable sharing of the benefits arising from biological resources. But the United States has not ratified it, and the 12 countries say that its principles are not being implemented. ■

## Cutbacks cost jobs at agricultural institute

K. S. Jayaraman, Hyderabad

Lack of funds is causing severe cutbacks at one of the few agricultural institutes to focus on the needs of the world’s poor.

About a quarter of the workforce at ICRISAT (the International Crops Research Institute for Semi-Arid Tropics) in Andhra Pradesh, India, will lose their jobs, as the institute attempts to make up for a shortfall in support from rich countries’ governments.

The institute will undergo restructuring and define a fresh research strategy aimed at securing new sources of funding, such as foundations and corporations, said director William Dar at a press conference earlier this month. This year’s budget of US\$22.3 million is down by \$1.5 million from last year. “We had no option but to streamline the functioning,” Dar said. The cut in staffing — the fourth in two years — will reduce numbers by 205 to 600, the lowest since the institute was established in 1972.

Institute spokesman Murli Sharma says that failure by donors to fulfil their commitments is not new. But this year, several donor governments have backed out, using the eco-



Dirt poor: women digging for groundnuts, an important crop in arid parts of India and Africa.

nomic slowdown caused by the 11 September attack on the United States as an “excuse”, Sharma says. He declined to identify the defaulters.

ICRISAT is one of 16 agricultural research labs run by the Consultative Group on International Agricultural Research, based in Washington DC. Until now, its main focus has been on the five mandated crops — sorghum, millet, groundnuts, chickpeas and pigeonpeas — that are important to poor farmers in arid regions of Africa and India.

Dar says the institute will now diversify into cash crops and livestock rearing, areas he hopes will attract new research sponsors.

These efforts to find new donors are starting to bear fruit. A New Delhi company, Proagro Seeds, has contributed \$300,000 for a three-year sponsored project on millet. And the Tata Trust in Mumbai has promised \$1 million for a three-year study on watershed management. “We need more non-traditional donors,” Dar says. ■

♦ [www.icrisat.org](http://www.icrisat.org)

## Cancer institute gives US breast-cancer policy the all-clear

**Washington** The United States is to continue its policy of recommending that women over 40 take regular mammograms to screen for breast cancer, the National Cancer Institute (NCI) announced last week.

The institute had been reconsidering the use of the technique in light of a review of the epidemiological evidence by researchers at the Nordic Cochrane Centre in Copenhagen, who questioned the benefits of screening last October (see *Lancet* **358**, 1340–1342; 2001). The researchers said that screening is not beneficial if deaths from cancer interventions such as chemotherapy are taken into account.

A panel of independent experts assembled to advise the NCI acknowledged that there are uncertainties about the benefits of mammograms, but ruled that there is enough evidence to continue recommending their use.

## British geologist rewarded for groundbreaking career

**London** This year's Crafoord Prize — a major award established to cover fields that are not represented by the Nobel prizes — has been awarded to Dan McKenzie, a geologist at the University of Cambridge, UK.

The Royal Swedish Academy of Sciences says that the US\$500,000 prize rewards McKenzie's contributions to the understanding of the dynamics of the lithosphere — the outer shell of the Earth — including plate tectonics, sedimentary basin formation and mantle melting.

McKenzie published several important papers in the late 1960s that helped to cement the idea of plate tectonics, the unifying — but at the time controversial — theory that explains the movements of continents and oceans. He went on to work on understanding and predicting earthquakes, and more recently helped NASA to analyse the surfaces and internal structures of Venus and Mars.

♦ [www.kva.se](http://www.kva.se)

## Web to reveal Pauling's methods and musings

**Washington** Digital versions of Linus Pauling's lab notebooks are to be released online on 28 February, the 101st anniversary of the double Nobel prizewinner's birth.

The 46 notebooks span the years between 1922 and 1994, and include details of the work that earned Pauling the 1954 Nobel Prize in Chemistry for his studies of the chemical bonds between atoms. Pauling campaigned against the development of nuclear weapons during the cold war, and in 1958 presented a petition to the United



**Web of life:** Linus Pauling's lab notes, which are to be published online, span a staggering 72 years.

Nations, signed by more than 9,000 scientists, that called for the end of nuclear testing. He was awarded the Nobel Peace Prize for this work in 1962.

The notebooks have been digitized by staff at Oregon State University, who say the records contain autobiographical musings as well as Pauling's experimental records.

♦ <http://osulibrary.orst.edu/specialcollections>

## Physicists call time on historic pendulum problem

**London** Researchers in the US state of Georgia have chimed in with fresh insight on a 350-year-old observation of how two linked clocks behave.

Christiaan Huygens, the Dutch physicist who invented the pendulum clock, noticed in 1655 that the pendulums of two clocks contained within the same housing fell into step so that they swung towards and away from each other in unison. Physicists have studied many such coupled oscillators since, but Huygens' observations have not been satisfactorily explained until now, researchers at the Georgia Institute of Technology say.

The team, led by physicist Kurt Wiesenfeld, reconstructed Huygens' apparatus from his lab notes, and modelled the system mathematically. Each pendulum affects the other through their shared housing. With a low-mass housing, the interaction is too great and one clock stops. Heavier housings weaken the link between the clocks, allowing differences in their frequencies to prevent synchronization. Huygens' observation, the researchers conclude, was the serendipitous result of using a housing of just the right mass.

## Grants aim to entice mums back to the lab

**Munich** The European Organization for Molecular Biology (EMBO) in Heidelberg, Germany, is offering fellowships to help re-integrate young scientists into research after having a baby.

The scheme, believed to be the first run

by a major research organization, is aimed in part at reducing the number of women who drop out of research. In 2001, a report by the European Commission found that up to 40% of female scientists leave the research system during the "leaky pipeline" between graduation and senior academic positions.

Eight two-year grants of 30,000 euros (US\$26,140) per year will be available as part of the EMBO plan. Until now, restart schemes have been restricted to a few funding organizations and universities in the United States, Britain and Switzerland.

♦ [www.embo.org](http://www.embo.org)

## Here's the catch — fishing could wreck Atlantic reefs

**London** Coral reefs in the northeast Atlantic Ocean are being irreparably damaged by unregulated deep-sea bottom trawling, marine researchers say.

Reefs in other parts of the world are endangered by rising ocean temperatures (see news story, page 947), but an examination of the coral structures off the coasts of Ireland, Scotland and Norway has found a different sort of damage: trawl scars up to 4 kilometres long in the 4,500-year-old reefs. In the most recent study, researchers from the United Kingdom, France and Norway found chunks of broken coral and associated sea life in the catches of trawler boats fishing off the west coast of Ireland (J. Hall-Spencer, V. Allain and J. Helge Fossa *Proc. R. Soc. Lond. B* **10.1098/rspb.2001.1910**; 2002).

"We urgently need improved management of offshore areas to protect ancient deep-water habitats and the fish they support," says the report's lead author, Jason Hall-Spencer of the University of Glasgow.



**Deep damage:** the effect of bottom-trawl fishing on Atlantic corals (top) can be devastating (above).



# Rebirth and regeneration

Japan's Center for Developmental Biology aims to become a world leader in regenerative medicine. It also wants to break with a system blamed for stifling the creativity of young researchers. David Cyranoski reports.



**R**emember Cumulina, the world's first cloned mouse? Even if you do, you might not remember the name of Teruhiko Wakayama, whose painstaking work made her birth possible.

In 1997, while a postdoc in Ryuzo Yanagimachi's lab at the University of Hawaii in Honolulu, Wakayama pioneered the technique for extracting nuclei from adult mouse cells and injecting them into eggs that underpinned the team's cloning success. His magic hands have since found employment at Rockefeller University in New York and Advanced Cell Technology of Worcester, Massachusetts, a biotech company that is famous for its cloning research.

But now, the 33-year-old Wakayama is going home to Japan to establish his own lab. He plans to investigate why cloning has such a low success rate, why many cloned animals die in the womb or shortly after birth, and how adult patterns of gene expression are reset by the cloning process. "I want to understand the

mechanisms of reprogramming," he says.

Wakayama's career history makes him an emblem of the institute he is joining, the Center for Developmental Biology (CDB) in Kobe. The CDB has been created to propel Japan to the forefront of the emerging field of regenerative medicine — the cocktail of stem-cell research, cloning and traditional developmental biology that aims to find ways of growing replacement tissues for those lost to age or disease.

## Creative drive

Significantly, the CDB has also placed a strong emphasis on youth and creativity, giving its group leaders autonomy to develop their own projects. In Japan, where established institute directors and professors typically hold sway, this is unusual. Indeed, many researchers blame the hierarchical culture of Japanese science for its failure to shine quite as brightly as the government's investment should allow — and for driving frustrated talent abroad.

In luring back a rising star such as Wakayama, then, the CDB has already made a point. "This is an exciting experiment to establish a new institute merging basic and medical developmental biology, as well as creating an environment that has been generally difficult to find in our universities," enthuses its director, Masatoshi Takeichi.

As well as having been established to develop regenerative therapies for Japan's ageing population, the CDB is part of an effort to regenerate the city of Kobe, which was devastated by an earthquake in 1995. Located on an island of reclaimed land in the city's harbour, the centre's four buildings cost some ¥7.1 billion (US\$53 million) to construct. They are



**Mouse proud:** Teruhiko Wakayama (left) and Ryuzo Yanagimachi show off their cloned animals.

packed with state-of-the-art equipment for cultivating and manipulating cells and tissues. The CDB's transgenic-mouse facility will house up to 150,000 experimental animals, and within 18 months the institute will gain a new bioinformatics centre. "It should be quite spectacular," says developmental geneticist Matthew Scott, who heads the multi-disciplinary Bio-X facility that is being launched at Stanford University in California.

But the CDB's ability to shed the stiff suit of the Japanese academic system will be crucial. "Japan has excellent scientists, but the current structure of the system does not always encourage independence in young investigators," says Cliff Tabin, a developmental biologist at Harvard University. "Consequently, their contributions have been less than what one might have expected."

The CDB will have a core research programme run by Takeichi and six other senior group directors. But it is to the 20-plus



**Free spirits:** centre director Masatoshi Takeichi wants group leaders to pursue their own interests.

STEPHAN MOTTESIER/AP

independent team leaders — of whom Wakayama is one — that the institute's senior managers are entrusting the centre's future. "We will give young researchers a degree of independence and use of facilities that is unknown in Japan," claims Yoshiki Sasai, one of the CDB's group directors. Averaging in their late thirties, team leaders will be given enough funding to hire three postdocs and two technicians. Collectively, their efforts have been dubbed the CDB's "creative-research-promoting programme".

The CDB has provided a unique opportunity for Wakayama, who says that his youth and the fact that he graduated from an unfashionable provincial university would normally rule out gaining an independent position in Japan. "The CDB evaluates you based on your work, rather than your university's name or your age," he says.

In establishing its *modus operandi*, the CDB has drawn on the experience of a previous effort to break the Japanese research mould: the Brain Science Institute (BSI) in Saitama, north of Tokyo, which was launched in 1997. Both are initiatives of RIKEN, Japan's Institute of Physical and Chemical Research — and, like the CDB, the BSI was established to give Japan a world-class centre in a hot scientific field. The BSI arranges its 36 teams into 11 groups, each with a director. But the team leaders all have great independence and interact without the rigid hierarchy that is typical of Japanese research centres.

This spring will be a decisive time for both institutes — the CDB will mark its official opening by holding a symposium in April, just as most of the BSI's research teams are up for their critical five-year review.

In putting its researchers on five-year contracts, and making the continuation of research teams dependent on evaluation, the BSI broke with Japanese tradition. One group has already been sent packing. "The review committee said it didn't need to exist," says

one BSI researcher. This is revolutionary in Japan, where harsh criticism is rare and most researchers stay in one place for life. "Reviewing must be like this, however heartless it may look, in order to maintain a high standard," says BSI director Masao Ito.

But unsurprisingly, many BSI lab heads are concerned about the imminent reviews, and worry that discarded researchers will struggle to find work. "I'm not sure Japan is ready for this. It's just not that fluid," says one.

### Testing times

The CDB faces an even more daunting review three years from now. As one of Japan's Millennium Projects — a ¥120.6-billion package set aside for science and technology initiatives — continuation after 2005 depends on a successful evaluation in that year.

Japan's current economic woes have also added to the pressure. The commercial application of research results is now a top priority for the Japanese government. Whereas the BSI merely had to establish itself as a major player in basic neuroscience — a task in which it has had reasonable success — the CDB is supposed to feed a 'biomedical cluster' that will commercialize its results and bring them to clinical fruition.

The CDB is connected by a 10-metre bridge to Kobe city's Institute of Biomedical Research and Innovation (IBRI), which opened last April and houses a hospital with 60 beds for clinical trials. The IBRI will share its reclaimed island with the Kobe International Business Center, which was launched



Phoenix from the flames: the construction of the Center for Developmental Biology (top) is part of the wider regeneration of Kobe in the wake of the city's devastation by an earthquake in 1995 (above).

last July to lure companies to the area. A medical business centre will also be opened within 18 months.

Nevertheless, Takeichi and his colleagues stress that basic research will remain at the core of the CDB's activities. "Our major goal is to understand how individual cells are organized to form a higher-order complexity," says stem-cell biologist Shinichi Nishikawa, another of the centre's group directors.

The CDB's research will range broadly across the fields of development and regeneration. Nishikawa, for instance, will study the processes by which stem cells renew themselves, differentiate to form more specific cell types, and die. Sasai's group will study frogs, chicks and mice to examine mechanisms of brain development. Other researchers will examine fundamental developmental processes such as segmentation in the vertebrate body plan. Takeichi completes the picture, being a world leader in studying how cells within tissues stick together using molecules called cadherins.

But perhaps the CDB's biggest challenge will be to generate an image of a truly international institute that is attractive to foreigners. In this, the BSI has been a pioneer, conducting meetings in English and boasting a research staff of whom a quarter are foreign.

The CDB has some way to go. Despite heavy recruitment advertising at last July's International Congress of Developmental Biology in Kyoto, chaired by Takeichi, the CDB has only been able to fill two slots, while negotiating a possible third, with foreigners.

Some also say that the CDB still has a way to go in reaching out to researchers elsewhere in Japan, noting the dominance of biologists from Kyoto University — from which Takeichi, Nishikawa and Sasai all hail. "They are far too Kyoto-based and cliquish," says one biologist from the Tokyo area, claiming that applicants from outside the region of Kyoto, Osaka and Kobe have not been dealt a fair hand. Takeichi rejects these charges, attributing the dominance of Kyoto to the university's strength in developmental biology.

Given the challenges faced by the CDB, Takeichi urges critics not to rush to judgement. Although RIKEN president Shunichi Kobayashi says it is ridiculous to suppose that the CDB will be shut down in 2005, the short time that the centre will have to justify its existence clearly weighs on Takeichi's mind.

"I'm worried that people will do short-term research and have many projects in parallel instead of concentrating on one in particular," he admits. "Our goal of creating regenerative medicine should be considered as a long-term project — 10 to 20 years. Evaluations should be made on this basis." ■

David Cyranoski is *Nature's* Asian Pacific correspondent.

Center for Developmental Biology

♦ [www.cdb.riken.go.jp/english/index.html](http://www.cdb.riken.go.jp/english/index.html)

Brain Science Institute

♦ [www.brain.riken.go.jp](http://www.brain.riken.go.jp)



# Voyage of the argonauts

A global network of sea-going floats is set to transform our understanding of the world's oceans. As the data start to roll in, researchers are lining up the problems they hope to solve. Rex Dalton reports.

**T**he cool, dense currents within the Labrador Sea were long thought to chart a hidden course down Canada's eastern seaboard. Hundreds of metres below the sea's surface, the cold waters were believed to slip southwards past Newfoundland before ducking under the warmer waters of the Gulf Stream and dispersing into the North Atlantic.

But four years ago, a group of oceanographers used data from 200 ocean-going floats to disrupt this neat picture. The floats were cast into the Labrador Sea and programmed to drift at different depths below its surface, rising every few days to transmit data to satellites overhead. To the surprise of the researchers, some of the floats took an unexpected course, travelling east around Greenland and across the Atlantic towards Iceland and Britain.

"It was odd," recalls Russ Davis, who ran the experiment from the Scripps Institution of Oceanography in La Jolla, California.



**A drop in the ocean:** water currents, temperature and salinity in remote areas are now being recorded by floats parachuted in for the job.



"Something very unusual and unexpected was happening." After further study, Davis and his colleagues incorporated this current into a whole new pattern for the circulation of the Labrador Sea (see left), one which has changed our view of the North Atlantic<sup>1</sup>. "It was probably the single most interesting discovery I've been involved with in my career," says Davis, reflecting on his 34 years of service.

## Data streams

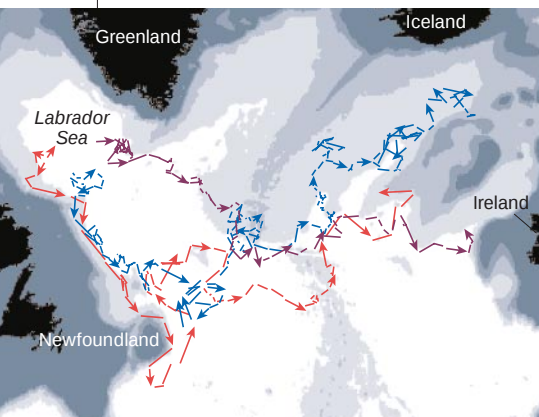
But the Labrador revelation could be just the first of many. Dean Roemmich, also at Scripps, estimates that over 300 data-collecting floats are now sending back temperature and salinity measurements from the world's oceans. Almost 700 more will be added over the next year. If the aims of this international endeavour, known as the Argo project, are realized, 3,000 floats will form a global network throughout the oceans by 2005.

Understanding ocean circulation is top of the project's agenda. Ocean currents move large volumes of warm and cold water

around the globe, so circulation has an important effect on climate. Researchers have mapped out broad circulation patterns, but much of the details of this movement, and the changes in temperature and salinity that accompany it, have not been filled in. "The ocean is the most poorly observed part of the climate system," says Sydney Levitus, director of the World Data Center for Oceanography in Silver Spring, Maryland.

Argo is the latest strand in the Global Ocean Observing System (GOOS), a decade-old drive by the Intergovernmental Oceanographic Commission, part of UNESCO, to put this right. The scheme aims to establish a permanent system for observing, modelling and analysing the world's oceans. The temperature of the ocean's surface, for example, can be measured by satellites, and instruments towed by ships can probe below the sea's surface. But by providing a regular flow of data from different depths and from across the globe, Argo will give the GOOS observation network a big boost. "For the first time," says Kensuke Takeuchi, an atmospheric physicist at Hokkaido University in Japan, "we will know the global-scale variation in the ocean in detail in real time."

Thirteen nations, including France, Britain, India, Japan and the United States, are now contributing floats to Argo, the total cost of which is likely to be between US\$30 million and \$40 million. The various participants have each designed different floats, but all the devices operate on similar principles. The ones developed at Scripps, for example, are cylinders around 150 centimetres long and 20 centimetres in diameter. Buoyancy is adjusted by pumping oil in and out of a bladder. This increases or decreases the float's volume, changing its density and caus-



**Flow chart:** data from floats in the Labrador Sea revealed completely new currents.



ing it to rise or fall through the water.

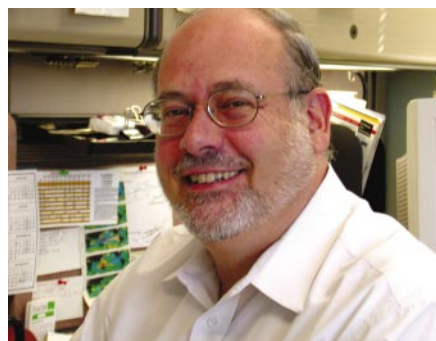
The float's ballast and bladder are tuned before launch so that it dives to a specific depth, which can be up to 2,000 metres. The US floats are programmed to surface and send data to satellites once every 10 days, and should be able to make up to 150 dives in their four- to five-year research lifetime. To minimize costs, floats are deployed by research vessels that happen to be going to or past target areas. This month, for example, an ocean-research vessel leaving Hobart, Australia, will deploy two floats in the Coral Sea and four in the Pacific Ocean west of Guam.

### Probe overboard

Engineers have also perfected a method for high-speed deployment, which can be used by crews of cargo vessels who are willing to throw the floats overboard, but who cannot stop to check on them if something goes wrong. The floats are contained in corrugated paper boxes that disintegrate rapidly in sea water. These boxes are also used to parachute the floats down to the sea from US Navy aircraft on training exercises.

Once deployed, the floats take temperature and salinity data as they move vertically through the water. Because the floats can only communicate when they are on the surface, newer generations capture data on the ascent rather than the descent to make the measurements more timely. Data from all the Argo floats are sent to three satellites owned by the US National Oceanic and Atmospheric Administration, and relayed to computers at the Brittany laboratories of IFREMER, the French marine research agency, and the Japan Meteorological Agency, based in Tokyo. From there, they can be freely accessed by interested scientists.

Climate modellers are particularly keen to get their hands on Argo data. These researchers often test their simulations by feeding them climate data taken, for example, in the 1950s. The models are then evaluated by comparing their predictions for the years between then and now with real measurements. But these tests are of limited value because data on the oceans are threadbare compared with those taken on land.



**Full stream ahead:** Sydney Levitus sees Argo as a major boost for climate research.

"Argo will fill in these gaps," says Detlef Quadfasel, an oceanographer at the Danish Center for Earth System Science in Copenhagen. "It will give a coverage similar to that which meteorologists have had on land for over 100 years."

Argo should also help to tackle more localized problems. Quadfasel set sail on 19 February for the Greenland Sea to investigate curious data produced by Argo floats in the region. One of the floats is trapped in an eddy, a column of water that spins round like a slow-moving whirlpool. Eddies at high latitudes interest oceanographers because they might act as pipes, channelling surface water down to lower levels of the ocean. Theoretical studies of eddies suggest that they should collapse after a few weeks or months<sup>2</sup>, but the Greenland Sea float has been trapped in its eddy since last March — hence Quadfasel's decision to study it first-hand.

Other groups plan to use Argo data to hunt for 'signals' of climate change — changes in the atmosphere or oceans that can be attributed to climate change rather than to normal variability. In 2000, a group of researchers from the Hadley Centre for Climate Prediction and Research in Bracknell, UK, predicted that some areas of the Indian Ocean should provide such a signal.

### Heat sinks

According to their model, climate change should significantly decrease the salinity, and increase the temperature, in water about 500–2,000 metres below the ocean's surface<sup>3</sup>. These changes would be caused by patches of surface water sinking to deeper levels. The signal arises because the variations in temperature and salinity caused by exposure to the atmosphere are more pronounced when compared with the surrounding deeper water, which has not been in contact with the surface.

The potential importance of this region led researchers at the Southampton Oceanography Centre, UK, to deploy some of Britain's Argo floats in the Indian Ocean. Two are already adrift, and researchers from the centre are setting sail this week to release 25 more. Eventually, they hope to have about 40 across the ocean. When the data start coming in, they should help to clarify how accurate the model is, and just how climate change will affect the ocean.

Researchers are already looking towards the next generation of equipment. Future floats will, for example, carry biochemical or physical instruments, such as turbulence sensors to help measure mixing between upper and lower regions of the ocean. And engineers at Scripps and the University of Washington in Seattle are developing floats that can be programmed to follow specific paths rather than drifting with the current. The Scripps prototype is designed to cruise



**Networking:** Dean Roemmich (right) and Russ Davis with one of the Argo floats.

for up to a year, covering around 7,000 kilometres in that time. "It's not quite ready for prime time," says Davis, "but it's close."

Argo researchers also hope to make use of the 66 low-orbiting Iridium satellites. Originally intended to provide a global mobile-phone network, they are now being used for US military communications after their original backers went bankrupt. By increasing the number of satellites available to the project, the Iridium system would cut the time the floats have to spend at the surface before they can send their data. And unlike the present system, the Iridium satellites would also let researchers transmit signals to the floats, allowing them to reprogram the devices mid-voyage.

With new floats in the pipeline and a mass of data waiting to be analysed, these are exciting times for the Argo project. After years of planning, the Earth's oceans will soon have a network of mobile observatories, albeit simple ones, that can begin to provide the kind of coverage that land-based centres currently offer. Those involved have no doubts about the project's impact. "Argo," says Takeuchi, "will open a new era for oceanography." ■

**Rex Dalton is Nature's West Coast US correspondent.**

1. Lavender, K. L., Davis, R. E. & Owens, W. B. *Nature* **407**, 66–69 (2000).
2. Marshall, J. & Schott, F. *Rev. Geophys.* **37**, 1–64 (1999).
3. Banks, H. T., Wood, R. A., Gregory, J. M., Johns, T. C. & Jones, G. S. *Geophys. Res. Lett.* **27**, 2961–2964 (2000).

Argo project

♦ [www.argo.ucsd.edu](http://www.argo.ucsd.edu)

Argo Information Center

♦ [argo.jcommops.org](http://argo.jcommops.org)

GOOS

♦ [ioc.unesco.org/goos](http://ioc.unesco.org/goos)

# Locking horns

Rival zoologists are sparring over some twisted horns from an Asian cow-like creature. Are the specimens the reality behind a Cambodian myth, or clever fakes by local artisans? John Whitfield sifts the evidence.

Is there now, or has there ever been, an animal called the spiral-horned ox? It's a question that leads to unexplored forests, ancient texts, eastern mythology — and to rival camps of zoologists, one contending that the world's museums contain genuine specimens of a mysterious cow-like ungulate, the other dismissing them as fakes.

The story began in 1994, when two researchers from Dresden in Germany described a new species from the Vietnam–Cambodia border<sup>1,2</sup>. The animal is still known only from its forehead and its horns, which have spiral tips and corrugations along their length. The specimen count stands at 21, mostly collected in the past decade from the region's shops and markets.

No zoologist has seen any more of *Pseudonovibos spiralis*, as the beast was named. But Cambodian folklore mentions a snake-eating cow with twisted horns, called the Khting Vor. Believers assert that *Pseudonovibos*, with its snake-like horns, gave birth to the myth<sup>3</sup>. A description in a Chinese encyclopedia from 1607 of a beast that sleeps while hanging from a tree by its horns — horns that can cure snakebite — might also have been inspired by *Pseudonovibos*<sup>4</sup>.

Confusing results from DNA analysis have suggested that the Khting Vor's abilities include shape-shifting. To a group of Austrian and German researchers, who took DNA from the Dresden specimens, *Pseudonovibos* seemed closely related to sheep and goats<sup>5</sup>. But last year, Russian researchers, using a different DNA sequence and different specimens, asserted it to be a type of buffalo<sup>6</sup>.

Alexandre Hassanin, a molecular systematist at Pierre and Marie Curie University in Paris, has a simpler explanation: *Pseudo-*

**Fact or fiction? Are these horns from the Vietnam–Cambodia border evidence of a new species — known from myths as a snake-eating cow — or are they fakes?**

*novibos* never existed. From his own DNA analysis, Hassanin concluded that the animal's apparent goatiness was down to laboratory contamination with chamois DNA<sup>7</sup>. Many of the horns show evidence of having been twisted and carved, Hassanin asserts<sup>8</sup>. And this month, he published the second of two papers<sup>9,10</sup> arguing that DNA sequences from *Pseudonovibos* are identical to domestic cow DNA. Case closed, claims Hassanin.

## Ox or hoax?

Robert Timm begs to differ. True, there are fakes, he says, but Timm asserts that the two sets of *Pseudonovibos* horns at his workplace in the University of Kansas Natural History Museum in Lawrence are genuine. American hunters collected these in Vietnam in 1929 (ref. 11), decades before the Dresden team described the beast from what Timm says really were fakes. "There's no doubt in my mind that the specimens I have here are real animals," he says. From X-ray studies, he argues that the Kansas horns got their twists and ridges naturally<sup>12</sup>.

Ronald Pine of the Field Museum in Chicago, who has seen the Kansas specimens, has some sympathy with that view. "An awful lot of people seem to regard it as having been established that this animal doesn't exist," he says. "I think that conclusion is premature."

But to Richard Melville, a retired financier who has spent more than 40 years exploring and studying Cambodia and reading memoirs from the country's colonial past, the believers' claims are "tortured conjecture".

Melville suspects that the horns were manufactured, and were ascribed their powers because of the centrality of snakes to Cambodian mythology, which mixes Hindu and Buddhist traditions<sup>12</sup>. Researchers were misled by their cultural ignorance, he claims. "Scientists are trained to look for the facts, not the myths or tricks."

The two camps came together in print to argue their cases last December<sup>12</sup>. The next step, says Timm, is to swap specimens; he is developing DNA tests for the Kansas horns, and is willing to give others access to them. But he does not anticipate closure soon. "We're not going to have this resolved in the next two or three years," says Timm.

Of course, the row could be settled tomorrow if a Khting Vor wandered out of the remote forests of Cambodia. But Timm does not expect to get so lucky: "I'm sure it's gone extinct — people would have found it."

**John Whitfield works in Nature's science writing team.**

1. Peter, W. P. & Feiler, A. *Faun. Abh. Staatliches Mus. Tierkd. Dres.* **19**, 247–253 (1994).
2. Peter, W. P. & Feiler, A. *Zool. Abh. Staatliches Mus. Tierkd. Dres.* **48**, 169–176 (1994).
3. Dioli, M. *J. Zool.* **241**, 527–531 (1997).
4. Macdonald, A. & Yang, L. *J. Zool.* **241**, 523–526 (1997).
5. Hammer, S. E. *et al. Naturwissenschaften* **86**, 279–280 (1999).
6. Kuznetsov, G. V. *et al. Naturwissenschaften* **88**, 123–125 (2001).
7. Hassanin, A. & Douzery, E. J. P. *Naturwissenschaften* **87**, 122–124 (2000).
8. Thomas, H., Seveau, A. & Hassanin, A. *C. R. Acad. Sci.* **324**, 81–86 (2001).
9. Hassanin, A. *et al. C. R. Acad. Sci.* **324**, 71–80 (2001).
10. Hassanin, A. *Naturwissenschaften* DOI 10.1007/s00114-001-0291-x (2002).
11. Timm, R. & Brandt, J. H. *J. Zool.* **253**, 157–166 (2001).
12. Brandt, J. H. *et al. J. Zool.* **255**, 437–444 (2001).



R. FRANCIS/ROBERT HARDING; HORNS: R. M. TIMM

A. HASSANIN



# Why impact factors don't work for taxonomy

Its long-term relevance, few specialists and lack of core journals put it outside ISI criteria.

Sir — In an earlier Correspondence<sup>1</sup>, I explained why impact factors are irrelevant for judging the quality of taxonomy. E. Garfield<sup>2</sup> requested quantitative arguments to support my hypothesis that taxonomy follows different “regularities” from most sciences, making the impact factors as calculated by the ISI (formerly the Institute of Scientific Information) inapplicable.

If ISI impact factors are to judge research meaningfully, as discussed in a recent News Feature<sup>3</sup>, they have to be a roughly accurate estimate of the real impact of publications. This requires that: (1) the impact of a paper is expressed by citations and the citation impact is positively correlated with the quality and/or relevance of the paper; (2) most cited and citing journals are considered (easily possible if Bradford's law of scattering applies to a field, as most cited papers are published in a few core journals); and (3) a paper gets most of its citations in the first few years after publication. These requirements are not met by taxonomy, for the following reasons.

First, the number of taxonomists is declining. Consequently, taxonomy does not follow the ‘exponential curve’ of most sciences. The old literature is not overwhelmed by an avalanche of new papers. The peak of species descriptions (a rough surrogate for relevant publications) was before 1900 for most groups, so the average age of references in taxonomic publications is much greater than those in other scientific disciplines.

I have analysed 2,091 references from seven randomly chosen, comprehensive taxonomic papers. The mean age of the references is 61 years, the median 36 years (details available direct from F.-T. K.), with 98.5% of cited papers being more than two years old. In shorter taxonomic journal papers, F. Köhler<sup>4</sup> found the average age of citations to be 46.7 years (s.d. 30.5 yr) and 41.8 years (s.d. 22.7 yr). It is therefore pointless to judge taxonomists according to the ISI method of analysing citations over the preceding two years.

Second, the relevance of descriptive publications in this field remains the same over time; original descriptions have to be referred to for ever, independent of the paper's quality. Outside taxonomy, referring to original descriptions becomes superfluous with time and/or revisions. Einstein's papers are not always cited for his theory, unlike the original descriptions of, say, *Escherichia coli* or *Drosophila melanogaster*. This obsolescence under-

estimates impact only for authors of well-studied species<sup>5</sup>. However, in taxonomy this obsolescence effect is negligible.

Third, for any group of organisms there are at best a handful of (or frequently no) extant specialists. Therefore the chance to become cited by colleagues is relatively rare compared with other fields. The number of taxonomists and consequently the number of publications is this low for one reason, which has nothing to do with need or quality: decision-makers are generally more enthusiastic about other fields. For these reasons, taxonomic papers have a long-term impact. Sometimes taxonomists have to wait a generation to be heavily cited. The number of citations of their empirical taxonomic publications depends on the number of taxonomists working on the same field and whether these colleagues publish in the few taxonomic journals covered by the *Science Citation Index (SCI)*. These things are a matter of luck.

Fourth, there are no core journals for general taxonomy. These exist for cladistics, biogeography, chemical systematics, and so on, but not for species descriptions, revisions of genera, identification keys or inventories. Where to publish depends on which museum the material is in; which institute or learned society the taxonomist is affiliated to; or which serial has the funding to accept long monographs. Only 27 (42%) of the 64 entomological journals covered by the *SCI Expanded* publish taxonomic papers, whereas 898 (82%) of the 1,100 serials held by the entomology

library of the Natural History Museum in London probably contain taxonomic information. Because of the less-developed relevance hierarchy of taxonomic journals, the low proportion covered by the ISI puts taxonomy at a disadvantage.

Finally, the most important impact of taxonomy is the usage of identification keys which enable non-taxonomists to identify and work with a group of organisms. The use of such keys is generally not documented in reference lists, hence a crucial impact of taxonomy is missed by citation analysis.

Consequently, taxonomy has no citation classics which the ISI would uncover. E. O. Wilson's citation classics aren't his taxonomic works, and R. Sokal's treatises on biometry and on the method of numerical taxonomy (not an empirical taxonomic work, but a methodological textbook on classification) have far more citations than his taxonomic work. The *SCI* is not an appropriate means to judge taxonomy because taxonomy does not meet its requirements for a meaningful judgement.

**Frank-Thorsten Krell**

Department of Entomology, The Natural History Museum, London SW7 5BD, UK

[f.krell@nhm.ac.uk](mailto:f.krell@nhm.ac.uk)

1. Krell, F.-T. *Nature* **405**, 507–508 (2000).
2. Garfield, E. *Nature* **413**, 107 (2001).
3. Adam, D. *Nature* **415**, 726–729 (2002).
4. Köhler, F. *Amateurwissenschaft: Entwicklung, Beschreibung und wissenschaftssoziologische Analyse am Beispiel der Koleopterologie* p. 195 (Diploma thesis: Sociology, Univ. Köln, 1988).
5. van der Velde, G. *Nature* **414**, 148 (2001).

## Physics gets physical

Sir — The readable and entertaining Words article “Coming to Terms” by J. L. Heilbron (*Nature* **415**, 585; 2002) on scientific nomenclature is amusing, but potentially misleading as far as it concerns high-energy physics. It may be true that the first flippant names such as ‘quark’ and ‘gluon’ were bestowed by US researchers, but they have monopolized neither the discoveries nor the follies.

The inspirationally named gluon — the elementary particle carrying the force that ‘glues’ quarks together — was discovered at Germany's high-energy accelerator centre, DESY, in Hamburg in 1979, and the more prosaically named ‘intermediate boson’ was discovered at CERN, the European laboratory for particle physics, in Geneva in 1983.

Less seriously, I plead guilty to coining TOE as a non-anatomical acronym for

Theory of Everything in an article that appeared in *Nature* (**323**, 595–598; 1986). As for Grand Unified Theories or GUTs, the term was coined by my CERN collaborators Andrzej Buras (Polish/German), Mary K. Gaillard (American/French), Dimitri Nanopoulos (Greek) and myself (British). But we did not have the ‘guts’ to put the acronym in our paper on this subject (*Nucl. Phys. B* **135**, 66; 1978) — instead, we weaselled out and used the equally non-anatomical GUM for Grand Unification Mass. To the best of my knowledge, GUTs were first spilled, metaphorically, by Dimitri Nanopoulos in a paper (*Harvard Preprint HUTP-78/A062*) published later in 1978.

As well as making discoveries, we Europeans can be just as facetious, jocular and capricious as our US friends.

**John Ellis**

Theoretical Physics Division, CERN, CH-1211 Geneva 23, Switzerland

# Triumph of the naturalist

A unique data set from a natural laboratory of evolution.

**The Birds of Northern Melanesia: Speciation, Ecology, and Biogeography**

by Ernst Mayr & Jared Diamond  
Oxford University Press: 2001. 492 pp.  
£45, \$55

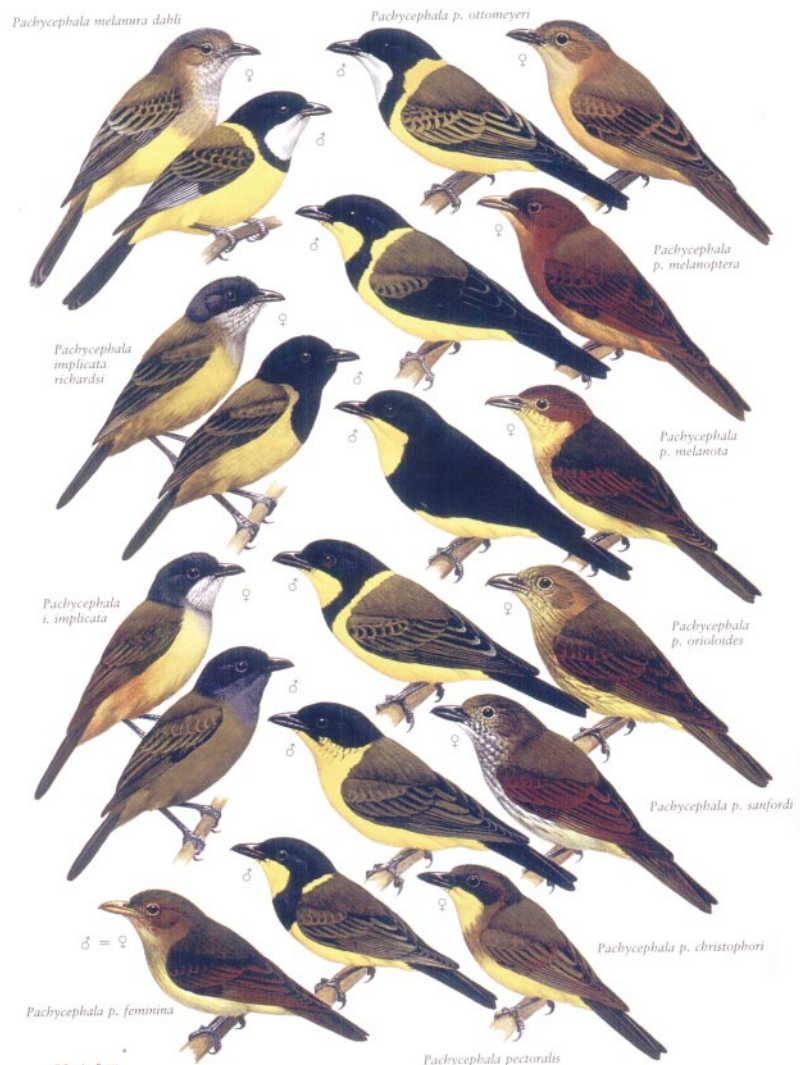
Stephen Pruett-Jones

East of New Guinea and northeast of Australia lies a collection of 200 islands in the Bismarck and Solomon archipelagos, collectively known as Northern Melanesia. This region of the Pacific is relatively unknown to the travelling public, but it has been extensively studied by ornithologists, who have conducted censuses of the islands for more than 150 years. Northern Melanesia supports 195 bird species, and among these are taxa illustrating every intermediate stage in the formation of new species, from undifferentiated recent arrivals on islands to endemic species and genera.

Northern Melanesia's usefulness as a natural experiment in avian speciation lies in the fact that all of the taxa dispersed there over water from just three sources: New Guinea, Australia and the Pacific islands to the east. Although many islands of Northern Melanesia have been joined to one another during past periods of lower sea levels, no island has ever been joined to either New Guinea or Australia. Thus, with just two exceptions, all avian speciation events in Northern Melanesia have occurred when colonizing populations were isolated by a water barrier. Furthermore, although all of the islands have similar climate and habitats, their physical areas span seven orders of magnitude — from just 0.002 to 35,742 square kilometres — and the number of islands occupied by an individual species ranges from 1 to 69.

As demonstrated by Ernst Mayr and Jared Diamond in *The Birds of Northern Melanesia*, these species have yielded the most complete data set ever compiled on the distribution and geographical variation of an entire avifauna on oceanic islands. Summarizing data from hundreds of biologists working in Northern Melanesia for more than a century, the authors completely revise the taxonomy of the region's birds, and include no fewer than 128 pages of tables, maps and appendices. Their book constitutes a compendium and analysis of data on the degree of taxonomic distinction of every species and every island population, their ecological and biogeographic attributes, and the sources of the colonists.

As optimistic as the authors are about their analyses, they are even more ebullient



**Plumed plurality: whistlers illustrate the great geographical variability of Northern Melanesia's birds.**

about the quality of their data: "We anticipate that much of the audience for this book may consist of biologists who are not interested in our conclusion, nor even in the questions that we ask, but who are instead interested in mining our raw data for their own purposes."

One of Mayr and Diamond's principal goals is to determine what this unique data set tells us about the process of speciation. These data can, indeed, tell us much. For example, although all species arrived in Northern Melanesia by dispersal, the likelihood of further differentiation is predicted by population size and dispersal ability: taxa with intermediate dispersal rates and moderate-to-large population sizes are most likely to show high endemism. Island size, distance from the mainland or other islands,

and habitat preferences also affect the degree of endemism of taxa. The data also highlight the role of hybridization in limiting differentiation. Northern Melanesia, for example, contains six cases of hybrid taxa showing traits of both of the parental taxa, and in most cases these hybrids arose when individuals of two different subspecies simultaneously colonized a previously uninhabited island.

The species complex of the golden whistler (*Pachycephala pectoralis*) illustrates both the extreme geographical variability of Northern Melanesian birds and the value of studying this area's avifauna. This species complex is centred on New Guinea, which supports seven sympatric species — that is, species occupying the same geographical area — many of them extremely similar in



appearance. *Pachycephala pectoralis*, which is widely distributed in Northern Melanesia, also lives sympatrically on some islands with another very similar species, *P. melanura*. But *P. pectoralis* has radiated within the islands of Northern Melanesia, forming at least 16 distinct subspecies that vary tremendously in plumage, song and ecology (and 66 subspecies across the whole of their range, from Java in the west to Samoa in the east). Whereas most subspecies in Northern Melanesia are sexually dimorphic, with males brighter in plumage than females, one subspecies (*P. p. feminina*) has monomorphic males and females, with both sharing the plumage characteristics of females of other populations. Furthermore, outside the Bismarck Archipelago is another subspecies with monomorphic sexes (*P. p. xanthoprocta*), but in this taxon both sexes share the bright male plumage. Numerous observations of hybridization confirm that these divergent, geographically separate populations are subspecies, but we do not understand why in some areas very similar whistlers fail to hybridize, and yet in others very disparate forms hybridize readily.

As the authors acknowledge, their taxonomic revision of the birds of Northern Melanesia, as well as their analyses of geographical variation, are presented without genetic sequence data or analyses of taxa with regard to phylogeny. It is unlikely, however, that such data will drastically change many of the authors' conclusions, particularly with respect to their analyses of island biogeography, but it will be important to use sequence data to confirm the evolutionary relationships of taxa presented here. In particular, sequence data will permit comparisons of genetic and morphological divergence in individual taxa. They will also allow more complete analyses of hybridization, and an examination of the connection between gene flow and morphological differentiation. Nevertheless, because few, if any, such studies of the avian taxa from this region are being carried out, there is, as the authors imply, no reason to wait for molecular data before examining geographical variation and speciation in these taxa.

Today, when the juggernaut of genomics and molecular genetics threatens to crush all other areas of evolutionary biology, the image of a naturalist venturing afield with binoculars and a notebook may seem little more than a quaint and outmoded view of how science is done. But, as demonstrated in this book, the combined data from such field biologists can, when supplemented with careful morphological studies, provide a unique and powerful data set with which to examine critical issues in ecology and evolution.

Stephen Pruett-Jones is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

## Sustaining tropical agriculture

### Exploring Agrodiversity

by Harold Brookfield  
Columbia University Press: 2001. 608 pp.  
\$75, £54

### Securing the Harvest: Biotechnology, Breeding and Seed Systems for African Crops

by J. DeVries & G. Toenniessen  
CABI Publishing: 2001. 224 pp.  
£27.50, \$50 (pbk)

Calestous Juma

Debates over the role of new technologies — especially genetic modification — continue to dominate international policy discussions. Advocates of biotechnology invoke images of the bountiful supplies of food that the technology can deliver. But critics stress the risks associated with the new technology and urge caution, or even outright bans, on its use. The din generated by this confusion of views has drowned out most efforts to present reasoned positions.

The absence of detailed scholarly works exploring these issues has conspired to perpetuate a false dichotomy between agricultural diversity and technological innovation. Two new books give the policy-making community a chance to go beyond the rhetoric and to assess the practical implications of biotechnology for sustainable agriculture.

In *Exploring Agrodiversity*, Harold Brookfield brings together a rich collection of evidence to underscore the importance of ecological diversity in the resilience of small-scale farming. His scepticism about the role of biotechnology stands in contrast to *Securing the Harvest*, which is a meticulous

attempt to identify the opportunities for using biotechnology to improve traditional agriculture. Whereas Brookfield seeks to celebrate the resilience of small-scale farming systems worldwide, DeVries and Toenniessen focus on ways in which they can be improved to expand the economic value of African agriculture.

The books are rooted in two different analytical traditions. Brookfield builds on his peerless work documenting diversity in farming systems around the world; DeVries and Toenniessen bring to their book unparalleled practical experience in agricultural technology, accumulated over the decades through the pioneering efforts of the Rockefeller Foundation.

Both books acknowledge the socioeconomic risks associated with previous agricultural models, but do so from different vantage points. Brookfield extends the critique of the green revolution and argues for a cautious approach that seeks to protect small-scale farmers. DeVries and Toenniessen, on the other hand, call for the use of biotechnology to improve small-scale African agriculture.

Brookfield acknowledges the possible contributions of transgenic agriculture, but suggests that small-scale farmers are likely to benefit more from "lower levels of technology". DeVries and Toenniessen contend that African agriculture "has yet to be transformed from subsistence, low-yield systems dependent on shifting cultivation to efficient, modern systems capable of producing regular surpluses". The two books appear to be presenting contradictory arguments, but a closer examination reveals that they are indeed complementary, offering new insights into the future of tropical agriculture.

Brookfield elegantly outlines the importance of agrodiversity. However, his data provide no compelling evidence that



What role might biotechnology play in helping Africa's small-scale farmers to increase their crop yields?

BETTY PRESS/PANOS PICTURES

biotechnology would necessarily undermine diversity. His use of analogies from the green revolution ignores the fact that farming systems can be redesigned to reflect new social goals. The book does not distinguish between the limited role that biotechnology can play in solving technical problems and the wider critique of corporate control of technology.

Fundamentally, the challenge lies in the choice of farming system, and this is independent of specific technologies. It is here that DeVries and Toenniessen offer complementary perspectives, especially through their emphasis on identifying technical opportunities for improving crops such as maize, sorghum, rice, cowpea, cassava and banana. Unlike Brookfield, they are concerned with socio-economic viability and offer new insights into the development of the seed sector. However, they pay less attention to ecological issues.

The two books use different units of analysis. Brookfield focuses on farming systems, whereas DeVries and Toenniessen are concerned with improving individual crops. By downplaying the role of technological innovation, Brookfield's analysis ignores the major challenges for African agriculture, such as its low productivity, and his approach tends to be over-optimistic about the ability of Africa's current systems to meet its growing food needs. DeVries and Toenniessen, however, leave themselves open to possible charges of technical determinism by not including a more detailed discussion of the policy implications of their recommendations. But read together, the two books offer unique insights into the relationships between ecosystem management and technological innovation.

They also offer divergent policy approaches. Brookfield makes a passionate appeal for global social justice as a way of protecting small-scale farmers from the risks of new technologies. But political appeals are not enough to protect farming systems from change. Nor will they be protected by diplomatic victories in international forums. DeVries and Toenniessen, on the other hand, pin their hopes on the co-evolution of technological change and institutional innovation, but they underestimate the challenges associated with these processes. And to have a dynamic agro-ecological system, all technological options must be kept open.

If read individually, the two books would only fuel the debate that is raging over agricultural biotechnology. But together, they provide clear indications on how to promote sustainable agriculture. ■

*Calestous Juma is in the Science, Technology and Innovation Program, Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA. He is a former executive secretary of the United Nations Convention on Biological Diversity.*

## Science in culture

### Spotlight on a visual language

**Do Piet Mondrian's beliefs about the aesthetic appeal of his art stand up to scientific scrutiny?**

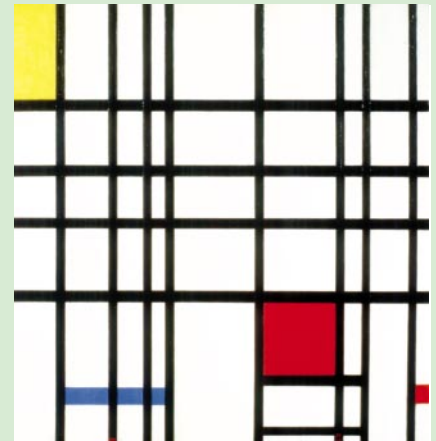
*Richard Taylor*

Piet Mondrian's abstract paintings are celebrated for their simple yet carefully arranged grids of strong black lines and enclosed rectangles of primary colours, which the painter considered to be reflections of the natural laws of the Universe. His paintings convey an impression of balanced harmony. But, as the audience at a visual-sciences conference was told last December, Mondrian's story is not as straightforward as the lines he painted.

Never before have the Great Masters of the art world received such scrutiny from beyond the traditional boundaries of art theory. Recently, painter David Hockney and optics scientist Charles Falco challenged five centuries of figurative art when they presented evidence that da Vinci, Raphael, Caravaggio, Van Eyck, Velázquez and Rembrandt may have traced images projected onto their canvases using lenses, prisms or mirrors (see *Nature* **412**, 860; 2001). To many, their investigation has relegated masterpieces of figurative art from displays of creative genius to the product of simple copying. So, what does science have to say about the abstract works of modern art?

Abstract art covers a vast visual spectrum, from Mondrian's geometrical patterns at one end to Jackson Pollock's intermingled swirls of paint at the other. Pollock's unorthodox style of dripping paint onto the canvas led to a popular belief that his 'compositions' were devoid of patterns. However, my analysis, with Adam P. Micolich and David Jonas, shows that his paintings are fractal, similar to nature's scenery (see *Nature* **399**, 422; 1999). Fractals consist of structure that repeats at different magnifications, building patterns of immense complexity. Unlike Pollock, Mondrian viewed nature's complexity with distaste, believing that it hid a purer form of natural order, a fundamental balance and equilibrium that "appears under a veil" of nature's erratic surface.

To capture this quality on canvas, Mondrian developed his simple visual 'language' of primary colours and straight lines. He maintained that the correct arrangement would deliver a profound aesthetic impact; sometimes he spent weeks deciding on the precise positioning of a single line. Remarkably, Australian artist Alan Lee told the conference audience that the theory collapses when put to the test. Lee created eight of his own paintings based on Mondrian's basic design elements of intersecting black vertical and horizontal lines enclosing coloured regions. However, he composed his patterns randomly. In visual-perception tests, 10 art experts and more than 100 non-experts were then presented with an array of 12 paintings and asked to identify the four of Mondrian's carefully composed patterns



**Mondrian's *Composition with Red, Yellow and Blue* (1939–42): the artist could spend weeks deciding on the positioning of a single line.**

and the eight random patterns produced by Lee. The results were the same as if the subjects had been blindfolded — the two types of pattern were indistinguishable.

Mondrian's obsession with the location of his lines extended to their angle. In one of the more notorious exchanges in modern art history, he argued fiercely when colleague Theo van Doesburg proposed that their visual vocabulary should be broadened to include diagonal lines. Mondrian passionately believed that the diagonal represented a disruptive element, and he threatened to break with the 'De Stijl' art movement that had formed around his aesthetic ideals. Branka Spehar, a perception psychologist from the University of New South Wales in Sydney, presented results that question Mondrian's belief. In her study, she showed 20 subjects images generated by tilting three of Mondrian's paintings at four orientations. The four orientations of each painting, including the one intended by Mondrian, were paired in all possible combinations and the subjects were asked to express a preference within each pair. The results, based on 72 trials for each subject, indicate that people show no aesthetic preference between the original orientations intended by Mondrian and the oblique ones.

Few will dispute the artistic value of Mondrian's visual language. But as scientists turn their expertise towards some of the world's most treasured paintings, their results can be unexpected. Like characters in a detective story, scientists are simply contributing their own unique clues to one of civilization's great questions — the meaning of art. ■

*Richard Taylor is in the Department of Physics, University of Oregon, Eugene, Oregon 97403, USA.*

*"The art of seeing and the seeing of art" was held at the Centre for Visual Sciences, Australian National University, Canberra, on 5–7 December 2001.*

♦ <http://cvss.anu.edu.au/artsci>



# The future of sex

R. John Aitken and  
Jennifer A. Marshall Graves

**T**he desperate plight of the human spermatozoon is clearly reflected by the poor fecundity of our species. Human spermatozoa stand apart from the gametes of virtually all other mammals in the paucity of their phenotype, the inadequacy of their function, and the sensitivity to fragmentation of their mitochondrial and nuclear genomes. Roughly one in seven Western couples seek treatment for infertility, mostly because of problems with semen quality.

Even when a human spermatozoon achieves the fertilization of an oocyte, damage can crop up in the next generation. All dominant mutations in our species (such as those that cause achondroplasia, multiple endocrine neoplasia and Apert's syndrome) seem to arise in the male germ line. Spermatozoa are also important mediators of the environmental contribution to cancer in young adults and children; for example, heavy smoking in fathers is reported to confer a fourfold increase in cancer risk to their children.

The impact of environmental toxicants and the innate inadequacy of human spermatozoa are compounded by the advent of effective contraception and the introduction of assisted-conception technologies. This lifting of the selection pressure on fertility means that those endowed with genes for high fecundity have lost their advantage over those without. As a result, future generations are bound to experience a further decline in semen quality and, ultimately, human fertility.

What mechanisms are responsible for the poor fertilizing potential and genetic damage

shown by human spermatozoa? Two main causes of germ-cell dysfunction have recently been discovered: gene deletions on the long arm of the male sex-determining Y chromosome, and oxidative stress. We believe that these aetiologies may be associated.

The Y chromosome is particularly vulnerable to gene deletions because it is not a matching partner for the X chromosome, so it cannot retrieve lost genetic information by homologous recombination. Over the past 300 million years, the mammalian Y chromosome has been reduced from a pairing partner to the X chromosome to a shadow of its former self, rescued only by a large addition from a non-sex-determining chromosome in 'placental' mammals. Many of the remaining genes have acquired functions essential for sex determination and spermatogenesis.

The original Y chromosome contained around 1,500 genes, but during the ensuing 300 million years all but about 50 were inactivated or lost. Overall, this gives an inactivation rate of five genes per million years. The presence of many genes that have lost their function (pseudogenes) on the Y chromosome indicates that this process of attrition is continuing, so that even these key genes will be lost. At the present rate of decay, the Y chromosome will self-destruct in around 10 million years. This has already occurred in the mole vole, in which the Y chromosome (together with all of its genes) has been completely lost from the genome.

Accelerated degeneration of the Y chromosome is found in the 5–15% of severely infertile men whose infertility is caused by wholesale deletions of parts of this chromosome. Because mutations that cause infertility cannot be inherited, the relative abundance of Y-chromosome deletions in male patients suggests an extremely high rate of spontaneous DNA damage. Even microdeletions on the Y chromosome destabilize its transmission, frequently causing it to be lost during gamete production.

One important mechanism by which DNA damage is induced in the male germ line is oxidative stress. Spermatozoa are particularly vulnerable to this because they generate reactive oxygen species and are rich in targets for oxidative attack. Moreover, because they are transcriptionally inactive and have little cytoplasm, spermatozoa are deficient in both antioxidants and DNA-repair systems. Oxidative stress is thus a major cause of male infertility, and contributes to the high rate of DNA fragmentation in spermatozoa.

Such DNA fragmentation probably predisposes the cell to mutagenic change, which would become fixed as a deletion in the embryo by aberrant recombination. The Y

## Human spermatozoa

*The vulnerability of the Y chromosome will be a key factor in shaping the evolutionary future of our species.*

chromosome is susceptible to such recombination because of its high frequency of repetitive elements. Fragmentation induced by free radicals in the Y chromosome's DNA might also cause other post-fertilization genetic changes, such as insertions and amplifications. Because mutations that originate in this way precede the embryo's first cleavage division, they will enter the germ line and contribute to infertility and morbidity, including cancer, in the offspring.

At present, we have no idea what causes oxidative stress, DNA fragmentation and functional incompetence in human spermatozoa. But we do know for certain that such events put pressure on the vulnerable Y chromosome. In the long term, absolute selection against males with deletions that confer sex reversal or sterility will create strong pressure either to retain (and amplify) fertility genes, or for any fertile variant that replaces it. Could the present race of humans eventually be replaced by a new variant (or several independent variants that cannot cross-hybridize) with an alternative sex-determining/differentiation system? Such a new hominid race could differ from present humans in many other characteristics, depending on the gene pool of the new variant's handful of founders.

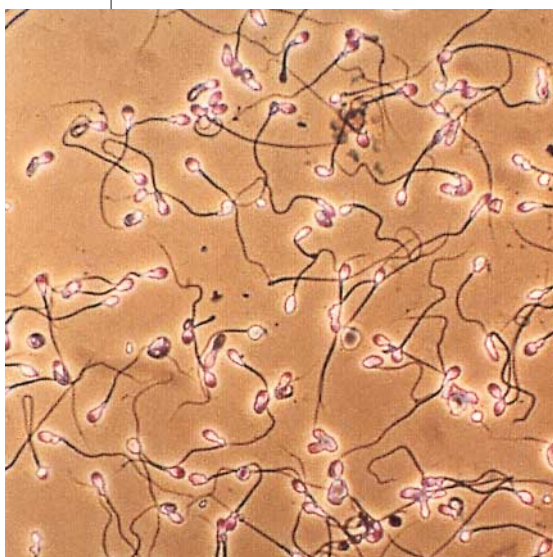
There is evidence of rapid 'selective sweeps' in the Y chromosome's evolution. These take place when a Y chromosome with an allele that confers a big selective advantage rapidly replaces other Y chromosomes in the population. As the Y chromosome is never broken up by recombination, whatever alleles lie at other genetic loci — even if they are deleterious — will 'hitch-hike' to fixation along with the advantageous variant. ■

R. John Aitken is at the Hunter Medical Research Institute and is in the Discipline of Biological Sciences, University of Newcastle, Callaghan, New South Wales 2308, Australia.

Jennifer A. Marshall Graves is in the Research School of Biological Sciences, Australian National University, Canberra, ACT 2601, Australia.

### FURTHER READING

Aitken, R. J. *J. Reprod. Fertil.* **115**, 1–7 (1999).  
Marshall Graves, J. A. *Biol. Reprod.* **63**, 667–676 (2000).  
Just, W. *et al. Nature Genet.* **11**, 117–118 (1995).  
Kuroda-Kawaguchi, T. *et al. Nature Genet.* **29**, 279–286 (2001).  
Kamp, C. *et al. Mol. Hum. Reprod.* **7**, 987–994 (2001).



**Swimming against the tide: sperm quality seems set to decline still further in future generations.**

# Magnetic moments at Jupiter

Thomas W. Hill

The coming together of two spacecraft near Jupiter provided a unique opportunity to investigate the giant planet's magnetic field — and the results, collected in this issue, are stunning.

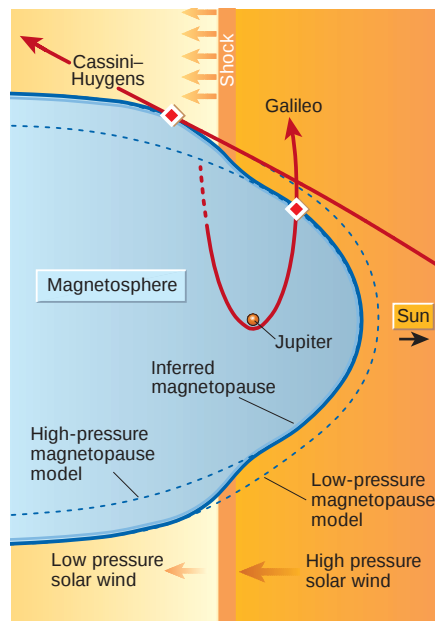
In January 2001, the Cassini–Huygens spacecraft rendezvoused with its sister craft Galileo in the vicinity of Jupiter. Planetary scientists seized the opportunity offered by this first-ever conjunction of two spacecraft at an outer planet, and were richly rewarded for their efforts. Observations by the two spacecraft were complemented by nearly simultaneous images from the Earth-orbiting Hubble Space Telescope and Chandra X-ray Observatory. As reported

in a series of seven papers in this issue<sup>1–7</sup> (beginning on page 985), this coordinated campaign has revealed several new clues to the complex dynamics of Jupiter's magnetosphere — the comet-shaped region of space that is filled by the planet's magnetic field.

Jupiter has by far the strongest magnetic field of any planet in the Solar System, and the most powerful magnetosphere. The Cassini–Huygens spacecraft flew within 135 Jupiter radii of the planet's centre (its equatorial radius is about 71,400 km). This distance of closest approach was fixed by the need to use Jupiter's gravity to boost the energy of the spacecraft just enough to guide it to its final destination, Saturn. Many scientists would have preferred a closer approach. Indeed, some (including myself) worried aloud that, at that distance, Cassini–Huygens would miss the magnetosphere altogether.

Jupiter's magnetosphere is easily the largest object in the Solar System, at around 20 solar diameters wide and several astronomical units in length (an astronomical unit is the mean radius of the Earth's orbit about the Sun,  $1.52 \times 10^8$  km). But the size of a magnetosphere varies, and although Jupiter's is immense, it is usually not quite that immense. Nature chose to be kind on this occasion, however: on 9–10 January 2001, the magnetosphere was unusually large — just large enough to reach out and touch the passing Cassini–Huygens spacecraft on two brief occasions, as Kurth *et al.*<sup>1</sup> report on page 991. In any case, the influence of Jupiter's magnetosphere is detectable far beyond its boundary surface (called the magnetopause) because it emits copious quantities of neutral gas, charged particles and electromagnetic radiation across the spectrum from radio to X-ray wavelengths.

The size of a magnetosphere is determined by the pressure balance at the magnetopause. The internal pressure is sustained by the planet's magnetic field and the plasma (hot ionized gas) trapped inside the field. The external pressure is provided by the solar wind — a fully ionized stream of plasma that flows continually, but variably, outwards from the Sun. The wind is highly supersonic, and its pressure variations tend to steepen into shock waves as they move away from the Sun. These interplanetary shocks produce sudden compressions and expansions of magnetospheres; when this



**Figure 1** The conjunction of the spacecraft Cassini–Huygens and Galileo at Jupiter. Jupiter's magnetic field is confined by the pressure of the solar wind to a region known as the magnetosphere (blue), bounded by the magnetopause. On 10 January 2001, both spacecraft encountered the magnetopause, within half an hour of each other. The dashed blue lines correspond to magnetopause equilibrium models for high and low solar-wind pressure. The near-simultaneous encounters do not fit a single model, but suggest a combination of the high- and low-pressure curves (solid blue line). Kurth *et al.*<sup>1</sup> conclude that the magnetopause was caught in the act of adjusting from a low-pressure to a high-pressure state, as an interplanetary shock wave passed between the two spacecraft positions. (Even though the solar wind typically travels at around  $400 \text{ km s}^{-1}$ , it would take 20 hours to cross this figure from right to left.)



**Figure 2** In orbit around Jupiter, the Galileo spacecraft caught this image of a volcanic eruption on Io, the innermost of Jupiter's largest moons. The plume extends to a height of 140 km above Io's surface.

happens in the Earth's magnetosphere, it causes disruptions of human technology, such as wireless communication, electric-power distribution and satellite navigation.

The two-spacecraft conjunction at Jupiter, coupled with a bit of good luck in the form of a passing interplanetary shock, allowed Kurth *et al.*<sup>1</sup> to catch Jupiter's magnetosphere in the act of compression. Although the pressure of the solar wind determines the size of the magnetosphere, no single model magnetopause can describe what the two spacecraft saw during their almost simultaneous magnetopause encounters (Fig. 1). Kurth *et al.* argue plausibly that the magnetosphere must have been in the process of compression by an interplanetary shock that had passed Galileo's position but had not yet arrived at Cassini–Huygens. Plasma measurements taken by Cassini–Huygens outside the magnetopause, just before and just after its immersion in the



## Box 1 Following the auroral footprints

As Cassini–Huygens flew past Jupiter, high-resolution images of the planet's aurorae were taken by the Hubble Space Telescope and the Chandra X-ray Observatory, which are both in orbit around the Earth. Like the northern and southern lights on Earth, Jupiter's aurorae are rapidly varying displays of coloured light emitted by atmospheric atoms that are electronically excited by the impact of charged particles from the magnetosphere. But the configurations of the aurorae, and the magnetospheric processes that generate them, are very different for the two planets.

Three of the four large galilean satellites — Io, Ganymede and Europa — produce distinctive auroral 'footprints' at Jupiter (page 997, Clarke *et al.*<sup>5</sup>). (The fourth satellite, Callisto, may also have a footprint, but if so it is lost in the bright glow of the higher-latitude auroral 'oval'.) Io's footprint is the largest and brightest, as befits this geologically active moon whose volcanic eruptions (Fig. 2) populate the magnetosphere

with heavy ions. Ganymede, which is the largest satellite in the Solar System, has the next-brightest footprint, probably because it has its own intrinsic magnetic field and carves out a small magnetosphere of its own within Jupiter's much larger one. These auroral footprints tell us about the electrodynamic interaction of Jupiter with its retinue of large satellites, and also provide points of reference for calibrating and fine-tuning our model of Jupiter's magnetic field.

At higher latitudes lies the bright, persistent auroral 'oval' (see Clarke *et al.*'s Fig. 1 on page 997), which superficially resembles Earth's auroral oval but arises from a very different process. Jupiter's auroral oval marks a circuit of electric current that enforces co-rotation of the magnetosphere with the planet. Isolated patches of auroral luminosity sometimes appear outside the main oval, and Mauk *et al.*<sup>6</sup> (page 1003) have found a link between one such patch, observed by the Hubble Space Telescope, and a distinctive plasma structure observed

simultaneously by the Galileo spacecraft at a distance of about 12 Jupiter radii from the planetary centre. The association is plausible, but the description of this event as 'Earth-like' is a subject of friendly debate.

The ultraviolet aurorae seen by the Hubble telescope account for most of the auroral power output. But the relatively few auroral photons emitted at much higher (X-ray) energies reveal unique clues about more energetic processes in Jupiter's magnetosphere. Chandra X-ray images (discussed by Gladstone *et al.*<sup>7</sup> on page 1000) show that these energetic photons emanate from higher latitudes on Jupiter than the main oval — not lower latitudes as previously thought. This result raises the intriguing possibility that the X-ray aurorae are produced not by magnetospheric heavy ions (whose density is very low at the large distances mapping to these high latitudes) but by heavy ions from the solar wind that have invaded the magnetosphere at its outer boundary.

T. W. H.

magnetosphere, are consistent with the passage of a shock in the interim.

The Cassini–Huygens spacecraft observed three such interplanetary shocks in the preceding two months, as it approached Jupiter from Earth. Gurnett *et al.*<sup>2</sup> (page 985) provide evidence that one of these shocks was probably responsible for a subsequent magnetopause crossing by Galileo. They also show that all three shocks were followed by distinctive brightenings of Jupiter's aurorae — analogous to the northern and southern lights on Earth — and associated radio emissions after an appropriate time delay. This result settles a long-running debate over the sign of the effect — whether the aurorae would brighten or diminish. Additional clues about Jupiter's aurorae were provided by observations from Earth-based telescopes that were timed to coincide with the spacecraft conjunction (see Box 1).

Jupiter's interaction with the solar wind is by no means passive; its magnetosphere leaves a clear imprint on the solar wind, even far upstream. Io is the innermost of Jupiter's

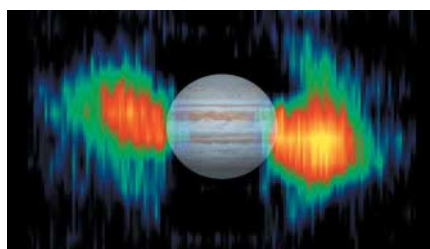


Figure 3 **Synchrotron emission around Jupiter, picked up by instruments aboard the Cassini–Huygens spacecraft, reveals electrons with unexpectedly high energies<sup>4</sup> — up to 50 million electron volts (yellow regions). An image of the planet constructed by the Galileo probe and the Hubble Space Telescope is superimposed.**

four largest moons — called the galilean satellites after Galileo who discovered them in 1610. Volcanoes on Io's surface supply heavy ions, mostly oxygen and sulphur, to Jupiter's magnetosphere (Fig. 2). There, the ions are energized and can escape their magnetic confinement if they are neutralized through charge-exchange interactions. These

energetic neutral atoms (ENAs) form a planetary wind that can propagate upstream against the solar wind (the atoms are not coupled to the wind because they are uncharged). Eventually the ENAs re-ionize and become part of the solar wind — but the wind, like the Sun, is mainly hydrogen, so the presence of the heavy ions is a distinctive signature of their origin in the volcanoes of Io. Cassini–Huygens has now confirmed the existence of the long-hypothesized ENA wind in both its neutral and re-ionized forms (Krimigis *et al.*<sup>3</sup> on page 994).

In addition to ENAs and charged particles, Jupiter's magnetosphere is also a prodigious source of non-thermal radio emissions. At frequencies below about 40 MHz, these emissions result from coupling processes between the magnetosphere and the ionosphere (the ionized level of Jupiter's atmosphere) that are associated with aurorae (see Box 1). Higher-frequency 'synchrotron' emission comes from relativistic (highly energetic) electrons trapped in the heart of Jupiter's radiation belt, only 0.5–3 Jupiter radii above the visible cloud level. The highest frequencies, corresponding to the greatest electron energies, are difficult to observe from Earth because they are lost in the thermal glow of the planetary disk. Taking advantage of the proximity of Cassini–Huygens, Bolton *et al.*<sup>4</sup> (page 987) used its radar instrument to separate the high-frequency synchrotron emission (at 14 GHz) from the thermal glow, and discovered an unexpected population of ultrarelativistic electrons with energies up to 50 million electron volts (Fig. 3).

The unique observations from the conjunction of two spacecraft are a fitting tribute to Galileo. In September 2003, after nearly eight years of successful orbital operations, the Galileo craft will be committed to the depths of Jupiter's atmosphere. But the work of Cassini–Huygens is just beginning. When it reaches Saturn in July 2004 the European Space Agency's Huygens Probe will separate from NASA's Cassini Orbiter and descend to the surface of Saturn's largest satellite, Titan, to investigate its intriguing, seemingly Earth-like atmosphere. Cassini will then begin an orbital tour, like that of Galileo at Jupiter, exploring Saturn's atmosphere, rings, satellites and magnetosphere. If its performance during the Jupiter flyby is any indication, Cassini–Huygens is ready for the task.

Thomas W. Hill is in the Department of Physics and Astronomy, Rice University, Houston, Texas 77251, USA.

e-mail: hill@rice.edu

1. Kurth, W. S. *et al.* *Nature* **415**, 991–994 (2002).
2. Gurnett, D. A. *et al.* *Nature* **415**, 985–987 (2002).
3. Krimigis, S. M. *et al.* *Nature* **415**, 994–996 (2002).
4. Bolton, S. J. *et al.* *Nature* **415**, 987–991 (2002).
5. Clarke, J. T. *et al.* *Nature* **415**, 997–1000 (2002).
6. Mauk, B. H. *et al.* *Nature* **415**, 1003–1005 (2002).
7. Gladstone, G. R. *et al.* *Nature* **415**, 1000–1003 (2002).

## Nuclear transplantation

## A monoclonal mouse?

Janet Rossant

One potential use of nuclear-transplantation — cloning — technology is to generate genetically matched tissues for treating adult patients. But there's a debate about whether mature adult cells are a good source of nuclei.

Behind all the fuss about the cloning of animals and prospects for cloning humans, some interesting scientific issues remain. For instance, can the nucleus of even the most specialized adult cell be completely 'reprogrammed', so that it can direct the development of a whole new embryo? Or are there subsets of cells in adult organisms — perhaps representing the unspecialized cells known as stem cells — that can be reset more easily? Hochedlinger and Jaenisch<sup>1</sup> tackle the question on page 1035 of this issue, and claim success in producing cloned mice by using nuclei from mature T or B cells. The results provide strong evidence that the nucleus of a differentiated cell can indeed be reprogrammed.

Given all the current hype, it might seem odd that the most commonly used approach to cloning — 'somatic-cell nuclear transfer' — was in fact developed some 40 years ago in amphibians purely to address such fundamental questions. The technique involves replacing the nucleus of an egg with that of a more specialized cell from an embryo or adult. The idea is that the egg's cytoplasm somehow reprogrammes the nucleus so that it is as developmentally versatile as that of a fertilized egg, and able to direct the formation of all the tissues that make up an organism — which would, of course, be genetically identical to the original cell.

These early experiments showed that nuclei from morphologically differentiated amphibian cells, such as tadpole intestinal epithelial cells<sup>2</sup> or adult skin cells<sup>3</sup>, could generate cloned tadpoles. But the success rate was low, and no adult frog was ever produced from an adult cell nucleus. By contrast, more recent experiments in mammals have shown that adult cell nuclei can produce viable clones<sup>4,5</sup>. But so few live-born offspring were obtained that it was possible that they did not come from mature cells, but from rare, less specialized adult cells such as stem cells.

So can highly differentiated adult cells be reprogrammed? To find out, some heritable mark is required that can unambiguously identify clones derived from a given differentiated cell. In mammals, the lymphocytes of the immune system are ideal for such an analysis. B lymphocytes are cells that produce soluble antibody molecules (immunoglobulins) in response to antigens (such as viruses or bacteria). Immunoglobulins are made up of a variable region, which recognizes antigens, and a constant region that

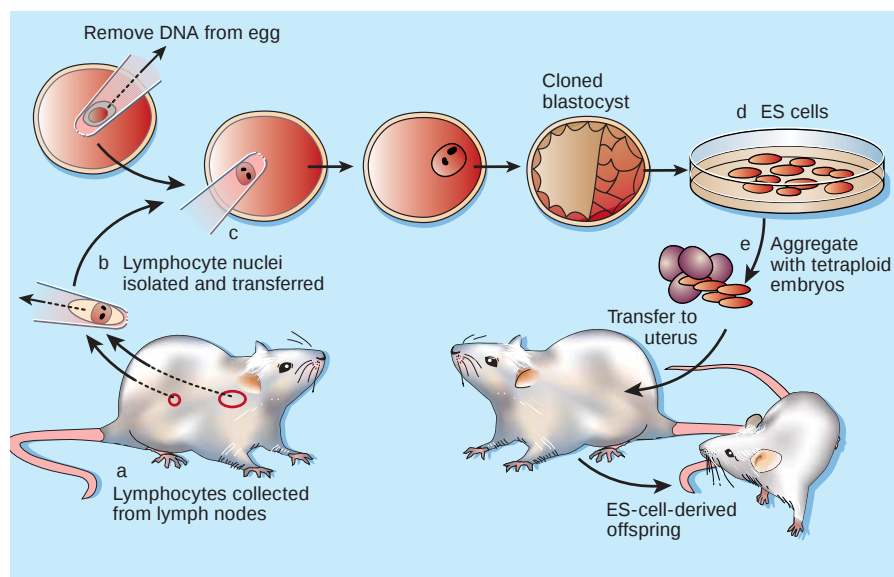
triggers immune responses. The genome encodes enough variable regions to respond to every type of antigen. But each mature B cell produces only one type of antibody, enabling it to respond to one antigen. To achieve this 'monoclonality', the cell rearranges its DNA to weld the variable region of choice with the constant region. Similarly, T lymphocytes respond to antigens through their antigen-receptor complex; this is also rearranged at the genomic level, so that each T cell expresses one type of receptor.

T and B cells are, therefore, rare examples of cells in which the genome sequence is altered as they mature. So if you use the DNA from a T cell or B cell for cloning by nuclear transfer, the genomic rearrangement should be detected in all the cells of the clones.

Unfortunately, lymphocyte nuclei seem to be rather recalcitrant to nuclear reprogramming. In Hochedlinger and Jaenisch's study<sup>1</sup>, the likelihood of a cloned mouse embryo reaching even the blastocyst stage —

a very early point in development — was only 4%. To get around this, the authors generated embryonic stem (ES) cells from the cloned blastocysts, and used a technique called tetraploid complementation<sup>6</sup> to produce mice from the ES-cell lines (Fig. 1). They made two such lines: 'LN1' had immunoglobulin gene rearrangements, whereas 'LN2' showed rearrangement of the T-cell antigen-receptor gene. So LN1 was derived from a B-lymphocyte nucleus and LN2 from a T-cell nucleus. From line LN1, Hochedlinger and Jaenisch generated viable mice that had rearranged immunoglobulin genes in all tissues. No live mice resulted from line LN2, although one dead fetus was delivered at term and a separate experiment showed that ES cells from line LN2 could contribute to several tissues, including the immune system, when combined with wild-type cells in chimaeras.

So does this experiment settle the question of whether nuclei from fully differentiated adult cells can be reprogrammed to allow normal development? The evidence is compelling, although the purists will say no. Using the ES-cell intermediates may have allowed time for extra reprogramming. Moreover, use of the tetraploid-complementation approach meant that the authors could not assess whether correct nuclear reprogramming could occur in placental tissues (see Fig. 1). Given that placental problems are one of the most consistent defects



**Figure 1** Hochedlinger and Jaenisch's technique<sup>1</sup> for generating cloned mice, using nuclei from mature immune cells. **a**, Lymphocytes were isolated from the lymph nodes, which consist almost entirely of mature B and T lymphocytes. These cells show rearrangements of immunoglobulin or T-cell antigen-receptor genes, respectively, whereas other tissues in the body have unrearranged genes. **b**, Nuclei were extracted from B or T cells and **c**, transferred into enucleated mouse eggs and cultured to the blastocyst stage of development. **d**, Embryonic stem (ES) cells were then generated from the cloned blastocysts and aggregated with tetraploid embryos (**e**), generated by electrofusion of two-cell embryos. This produces chimaeras in which the entire fetus is derived from ES cells but the placental structures are provided by the tetraploid cells. Resulting offspring are entirely ES-cell-derived and, if developed from successful transfer of a B- or T-lymphocyte nucleus, will show the appropriate gene rearrangements in all tissues.



in cloned mammals, this is a valid concern. Finally, the number of live cloned offspring produced per number of nuclear transfers was pitifully small — approaching one in a thousand. So we are still left wondering whether the nuclei of differentiated cells can only be reprogrammed under exceptional circumstances.

Why do nuclei from different adult tissues vary in the efficiency with which they can be reprogrammed? Could it be due to whether or not the tissues contain stem cells? That remains to be seen. The idea that stem-cell nuclei may be more easily reprogrammed came from using ES cells — which are even more developmentally versatile than adult stem cells — as a source of nuclei for nuclear transfer<sup>7,8</sup>. These studies showed that more cloned blastocysts developed to the point of live birth when the source of nuclei was ES cells rather than adult cells.

But fewer embryos produced from ES-cell nuclei reached the blastocyst stage, meaning that the overall percentages of live offspring per nuclear transfer were not very different. One problem with ES-cell-derived clones is that their expression of 'imprinted' genes — genes that are specifically expressed from either the maternally derived or the paternally derived chromosome, but not from both — is often abnormal. By contrast, a study<sup>9</sup> of cloned mice derived from adult cell nuclei showed that several imprinted genes were expressed normally. So stem cells from fetal or adult tissues (see, for example, refs 10, 11) might be a better choice than ES cells for testing the reprogramming potential of stem-cell nuclei.

These questions are of fundamental interest, but of course have practical implications. One of the potential uses of cloning is

in treating human diseases; the idea is to use some of a patient's nuclei to produce genetically identical early embryos, from which ES cells would be generated and used to grow healthy replacement tissues *in vitro*. But unless there is a real breakthrough in finding a source of adult nuclei that can be efficiently reprogrammed, all the talk about this 'therapeutic cloning' will come to nothing. In the largest study in mice to date<sup>12</sup>, only 35 ES-cell lines were generated from over a thousand nuclear transfers (an efficiency of just 3.4%). This will not be acceptable in humans, where eggs will be hard to come by. Reprogramming adult cells directly, without an oocyte intermediate, would seem a more viable alternative. Surely the time has come for the cloners to turn their attention to the molecular mechanisms of nuclear reprogramming in the egg, and to use the information to enhance the potential of adult cells for use in cell-based therapies. ■

Janet Rossant is at the Samuel Lunenfeld Research Institute and the Department of Molecular and Medical Genetics, University of Toronto, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada.

e-mail: rossant@mshri.on.ca

1. Hochedlinger, K. & Jaenisch, R. *Nature* 415, 1035–1038 (2002); online 10 February 2002 (10.1038/nature718).
2. Gurdon, J. B. *J. Embryol. Exp. Morphol.* 10, 622–640 (1962).
3. Gurdon, J. B. & Laskey, R. A. *J. Embryol. Exp. Morphol.* 24, 227–248 (1970).
4. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. *Nature* 385, 810–813 (1997).
5. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. *Nature* 394, 369–374 (1998).
6. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. *Proc. Natl Acad. Sci. USA* 90, 8424–8428 (1993).
7. Humpherys, D. et al. *Science* 293, 95–97 (2001).
8. Rideout, W. M. III et al. *Nature Genet.* 24, 109–110 (2000).
9. Inoue, K. et al. *Science* 295, 297 (2002).
10. Goodell, M. A. et al. *Nature Med.* 3, 1337–1345 (1997).
11. Toma, J. G. et al. *Nature Cell Biol.* 3, 778–784 (2001).
12. Wakayama, T. et al. *Science* 292, 740–743 (2001).

## High-energy physics

# The mass question

Edward Witten

Do the elementary particles known as neutrinos have mass? Yes, according to recent experiments. But how much? A surprising — and controversial — result suggests that the answer is not what we thought.

Neutrinos were long believed to be, like photons, massless particles that always travel at the speed of light. In the past few years, by studying neutrinos emitted by the Sun or created by cosmic rays in the Earth's atmosphere, physicists have learned that neutrinos actually have tiny but non-zero masses, roughly ten million times smaller than the mass of an electron. These masses are believed to result from physical processes occurring at energies well beyond those of known particle interactions. In *Modern Physics Letters A*, Klapdor-Kleingrothaus and colleagues<sup>1</sup> now claim to have observed

a new type of nuclear decay process. If this somewhat controversial finding holds up, it implies that the three types of neutrino have almost the same mass, and gives us a window on physics that goes far beyond our present knowledge.

To put the mass of the neutrino in context, consider the mass of other elementary particles. The electron, for example, is about 1,800 times lighter than the proton or neutron, and about 200,000 times lighter than the heaviest known elementary particles, which are the W and Z bosons and the top quark (Fig. 1). Why these masses vary



## 100 YEARS AGO

In reviewing my child's book, "Beautiful Birds", F. E. B., writing in your columns, says, "Why should he select the 'beautiful birds' only, and, by implication, condone the massacre of birds that have not that advantage?" The question is a misstatement of fact, which I hope you will allow me to show, though I can only do so by quoting myself. On the last page — which I daresay F. E. B. did not get to — there is this:

"Mother, promise not to wear any feathers except the beautiful ostrich feathers that you look so lovely in?" As soon as she promised, then all the beautiful birds in the world (and that means all the birds, for all birds are beautiful) will be saved," &c. (The italics are mine). This is the final promise and the goal to which I have been leading. May I ask F. E. B. whether, if he wished to arouse a child's interest and sympathies in any subject, he would choose the more or less salient material to do it with? Edmund Selous

I would commend to Mr. Selous Dr. Samuel Johnson's sound remark concerning a quite analogous statement. An orchard, observed the Doctor, would be properly described as barren of fruit, even if subsequent research discovered a dozen apples and pears upon two or three trees. Now Mr. Selous' book is called "Beautiful Birds." It is not called "Birds." It is clear, too, what Mr. Selous means by "beautiful." His plates and the greater part of his descriptions deal with the Paradisæidæ, Humming Birds, and other birds which everyone calls beautiful. I do not find chapter after chapter relating to partridges, quails, sparrows, and other "plain" birds. F. E. B. From *Nature* 27 February 1902.

## 50 YEARS AGO

In *Nature* of January 19, p. 92, a translation was published of resolutions passed at a conference held in Moscow last June on the theory of chemical structure in organic chemistry. It was stated there that "The Conference has clearly demonstrated the soundness of the theory of the structure of organic compounds due to the great Russian scientist, A. M. Butlerov; this theory lies at the basis of the whole of modern organic chemistry". The theory of resonance or mesomerism was said to be "directly opposed to the basic thesis of Butlerov's theory", and it was condemned as physically untenable and sterile. Such sweeping claims require examination.

From *Nature* 1 March 1952.

| Matter             |                                           |              |                | Force             |                         |
|--------------------|-------------------------------------------|--------------|----------------|-------------------|-------------------------|
| Leptons            |                                           | Quarks       |                | Bosons            |                         |
| Electron<br>0.0005 | Electron neutrino<br>< $2 \times 10^{-9}$ | Up<br>0.003  | Down<br>0.006  | Photon<br>0       | (Electromagnetic force) |
| Muon<br>0.106      | Muon neutrino<br>< $2 \times 10^{-9}$     | Charm<br>1.3 | Strange<br>0.1 | W, Z<br>80.4/91.2 | (Weak force)            |
| Tau<br>1.777       | Tau neutrino<br>< $2 \times 10^{-9}$      | Top<br>175   | Bottom<br>4.3  | Gluon<br>0        | (Strong force)          |

**Figure 1 The standard model of high-energy physics: fundamental particles and their masses (in GeV  $c^{-1}$ , where  $c$  is the speed of light). Leptons and quarks interact through exchange of the particles associated with three forces (weak, strong and electromagnetic) to form the matter we see around us. The fourth fundamental force, gravity, cannot yet be described within the framework of the standard model. Although we do not yet understand why, the matter particles form three 'families' in order of increasing mass. The observation of a rare nuclear decay by Klapdor-Kleingrothaus *et al.*<sup>1</sup> suggests that neutrino masses may not follow this trend, but are in fact similar in value.**

so much is a mystery, even in the modern standard model of elementary particles. By contrast, until recently, neutrinos seemed to be massless, and in the 1950s physicists thought they had worked out why.

The key is chirality. In biochemistry, chirality describes the 'handedness' of a molecule, which may look different from its mirror image. A simple molecule such as H<sub>2</sub>O looks the same as its mirror image, but a more complex molecule such as dextrose may not. That certain chiral molecules are important in biology, and their mirror-image molecules are not, is believed to reflect accidents in the evolution of life, rather than any inherent difference between the molecules.

Neutrinos have a similar kind of chirality. Elementary particles have an intrinsic quantum-mechanical 'spin'. Most particles can spin in a right-handed or left-handed sense around their direction of motion, but neutrinos always spin in a left-handed sense (Fig. 2). Like chirality in biology, this property may conceivably have its origins in a chance event, in this case an accident of the Big Bang. Such an intrinsic chirality is impossible for particles with mass (because the direction of spin of a massive particle can be changed by rotating the particle in its rest frame), so physicists concluded that neutrinos must have zero mass.

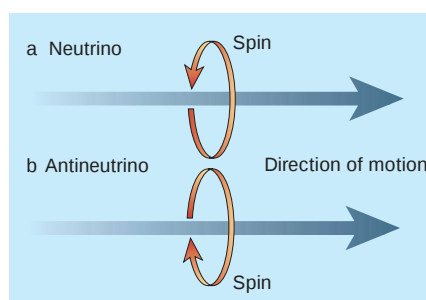
But there is a problem with this argument, and it has to do with antimatter. Every particle of elementary matter has a corresponding antiparticle, with the same mass but opposite electric charge. For example, the antiparticle of the electron,  $e^{-}$ , is the positron, denoted  $e^{+}$ . Similarly, the neutrino has an antiparticle: the antineutrino. The antineutrino has the opposite chirality to the neutrino — it always spins in a right-handed sense around its direction of motion (Fig. 2).

Apart from their chirality, how can you tell a neutrino from an antineutrino? They are both electrically neutral, so we cannot

distinguish them by their electric charge. But there is another apparently conserved charge in interactions between elementary particles: the lepton number. The electron and the neutrino are leptons, and the positron and the antineutrino are antileptons. The number of leptons minus the number of antileptons in an interaction is called the lepton number. Leptons and antileptons can be created by many processes, such as the decay of a neutron to a proton, an electron and an antineutrino. In this example, there are no leptons at the outset (the neutron is a 'baryon'), then one lepton (the electron) and one antilepton (the antineutrino) are created, so the lepton number does not change. Indeed, it is conserved in all the usual elementary particle processes.

The concept of lepton-number conservation was derived from experiment, and originally had no theoretical explanation behind it. In the 1970s, the newly developed standard model of high-energy physics offered some insight: given the particles assumed to exist in the standard model and the rules by which it is constructed, it is actually impossible to violate lepton-number conservation.

The standard model was barely in place



**Figure 2 Chirality is the spice of life. a, The neutrino spins in a left-handed sense around its direction of motion; b, the antineutrino spins in a right-handed sense.**

before physicists started trying to go beyond it. They wanted to build a unified theory that would motivate the existence of elementary particles and forces, rather than just describing them, as the standard model does<sup>2,3</sup>. In this more ambitious framework — optimistically dubbed 'grand unification' — lepton-number conservation is not automatic. Thus, a new perspective emerged<sup>4-6</sup>: lepton number should be very nearly conserved in nature because it is exactly conserved in the well-tested standard model; but it should be very slightly violated by the effects of grand unification.

If lepton number is not conserved, it no longer provides a way of distinguishing a neutrino from an antineutrino. They could, in fact, be two forms of the same particle. This particle has one state that spins one way and another state that spins the other way (Fig. 2), just like a particle with mass, such as the electron. So if lepton number is not conserved, neutrinos could have mass. But this mass can only be very small, because it arises from effects that are absent in the standard model. Direct measurement of such a small mass is difficult, but studies of the decay of the tritium nucleus have demonstrated<sup>7</sup> that one type of neutrino is lighter than about 2 electron volts.

A more subtle way of looking for neutrino mass depends on the fact that there are three kinds of neutrino: the electron neutrino, the muon neutrino and the tau neutrino (which are typically produced alongside electrons, muons and tau leptons, respectively). This leads to the possibility of an interesting quantum-mechanical effect: while travelling through a vacuum, one type of neutrino can convert spontaneously into another. This is known as neutrino oscillation, and can only happen if neutrinos have mass.

There is now extensive evidence for neutrino oscillations, both from neutrinos produced by cosmic rays in the Earth's atmosphere<sup>8,9</sup> and from neutrinos produced by the Sun<sup>10</sup>. (The interpretation in terms of neutrino oscillations has resolved a long-standing discrepancy<sup>11</sup> between the number of neutrinos expected from the Sun and the number we actually detect.) In this fast-moving area, experiment is well ahead of theory, and many important measurements are expected in the next few years. The results so far support the rough range of possible neutrino masses that arises from grand-unification theory. The experiments have also turned up a surprise: the measured 'mixing angles' (which determine the probability that neutrinos oscillate from one type to another) are much larger than theorists generally expected.

It seems logical to suspect that neutrino mass results from the non-conservation of lepton number. But the neutrino-oscillation measurements alone do not show that lepton number is not conserved. So can we do this



in some other way? This is what Klapdor-Kleingrothaus *et al.*<sup>1</sup> claim to have done, by observing the nuclear decay  $^{76}\text{Ge} \rightarrow ^{76}\text{Se} + 2e^-$ . This reaction is called neutrinoless double- $\beta$ -decay, as the final state contains two electrons (historically known as  $\beta$ -particles) and no antineutrinos — so the reaction violates the conservation of lepton number by two units. Taken together with the oscillation measurements, and assuming that the only relevant particles are the three known types of neutrino, the new result implies that the three neutrinos have approximately equal masses, probably a few tenths of an electron volt. This is a surprising result because other particle families, such as quarks and the charged leptons, do not have approximately equal masses (Fig. 1), and it will put a severe constraint on theories of the origin of neutrino masses.

Some caution is called for, however, because of the exceptionally difficult nature of the experiment. Criticisms of the assumptions made by the authors in analysing the background and extracting an extremely small signal have already been offered<sup>12,13</sup>. At any rate, planned future experiments using much larger quantities of  $^{76}\text{Ge}$  (or similar nuclei) will achieve much greater sensitivity. By extrapolating from the oscillation measurements, many physicists have guessed,

prior to this claim, that a sensitivity  $10^3$  or  $10^4$  times greater than that of this experiment may be needed to conclusively observe the violation of lepton-number conservation. Such sensitivity suggests how difficult, as well as how potentially rewarding, future experiments are likely to be. ■

Edward Witten is at the School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540, USA.  
e-mail: witten@ias.edu

1. Klapdor-Kleingrothaus, H. V., Dietz, A., Harney, H. L. & Krivosheina, I. V. *Mod. Phys. Lett. A* **16**, 2409–2420 (2001).
2. Pati, J. & Salam, A. *Phys. Rev. D* **10**, 275–289 (1974).
3. Georgi, H. & Glashow, S. *Phys. Rev. Lett.* **32**, 438–441 (1974).
4. Yanagida, T. in *Proc. Workshop on Unified Theory and Baryon Number in the Universe* (eds Sawada, O. & Sugamoto, A.) 95–98 (KEK, Tsukuba, 1979).
5. Gell-Mann, M. *et al.* in *Supergravity* (eds van Nieuwenhuysen, P. & Freedman, D. Z.) 315–321 (North-Holland, Amsterdam, 1979).
6. Weinberg, S. in *First Workshop on Grand Unification* (eds Frampton, P., Glashow, S. L. & Yildiz, A.) 347–362 (Math. Sci. Press, Brookline, MA, 1980).
7. Bonn, J. *et al.* (MAINZ collaboration) *Nucl. Phys. Proc. Suppl.* **91**, 273–279 (2001).
8. Fukuda, S. *et al.* (SuperKamiokande collaboration) *Phys. Rev. Lett.* **85**, 3999–4003 (2000).
9. Giacomelli, G. & Giorgini, M. (MACRO collaboration) Preprint hep-ex/0110021 (2001); <http://xxx.lanl.gov>
10. <http://www.sns.ias.edu/~jnb/SNexperiments/experiments.html>
11. Bahcall, J. N. & Davis, R. Jr in *Essays in Nuclear Astrophysics* (eds Barnes, C. A., Clayton, D. D. & Schramm, D. N.) 243–285 (Cambridge Univ. Press, 1982).
12. Ferruglio, F., Strumia, A. & Vissani, F. Preprint hep-ph/0201291 (2002); <http://xxx.lanl.gov>
13. Aalseth, C. E. *et al.* Preprint hep-ex/0202018 (2002); <http://xxx.lanl.gov>

## Biomechanics

# Walking with tyrannosaurs

Andrew A. Biewener

*Tyrannosaurus* terrorized the Earth — at least in the Hollywood version of history. But an estimate of the muscle volume in its hind legs suggests that the mighty giant could only walk, not run.

Over the course of history, vertebrates have evolved an enormous range of sizes, spanning well over six orders of magnitude in body mass. The largest and most captivating terrestrial giants were the dinosaurs, and *Tyrannosaurus* — although not the largest at around 6,000 kg — is perhaps the most famous and terrifying representative of this group. Some workers<sup>1,2</sup> have argued that bipedal tyrannosaurs and other huge dinosaurs could not move fast because their size would have imposed severe constraints on physiological and mechanical functions. But others claim that these creatures were much more athletic<sup>3,4</sup>.

An obvious difficulty in resolving this argument is that dinosaurs have been extinct for a long time, so reconstructing how they moved is a challenge. But on page 1018 of this issue<sup>5</sup>, Hutchinson and Garcia introduce a new biomechanical approach to the problem, applying an analysis of living animals to their ancient dinosaur relative. They show that *Tyrannosaurus* simply did not have large

enough leg muscles to produce the forces required for an animal of such size to run.

The skeletal muscles in all animals are made of the same contractile proteins, so their intrinsic capacity for generating force is very nearly the same. The force that can be produced depends on the cross-sectional area of a muscle's fibres. But as body size increases, the geometrical effects of scale mean that muscle capability does not increase proportionately. The force that a muscle can generate increases less rapidly than body weight, so, despite their greater volume, the muscles of larger animals generate less force per unit weight.

In addition, the ability of an animal's skeleton to support mechanical loads decreases with size because bone area does not increase nearly as fast as an animal's weight. Living terrestrial mammals can accommodate these problems of scale by altering their limb posture when they run: larger animals run on more erect limbs than much smaller animals, which gives their muscles greater

mechanical advantage<sup>6</sup> and allows them to maintain similar capacities of force generation and bone loading. But this only applies to animals as large as 300 kg or so. Above this weight, further changes in muscle mechanical advantage are probably limited<sup>7</sup>, and sustaining force capacity for movement at greater speeds becomes a problem.

So how fast might a 6,000-kg dinosaur have moved? Previous estimates of the speed and locomotive capacity of dinosaurs and other extinct animals have been purely qualitative. Some models are based on the limb motion deduced from the step length and stride frequency derived from fossilized tracks<sup>1,2,4,8</sup>. However, such estimates depend on assumptions about body mass distribution, limb posture and limb length, and about kinematic similarities between species. The data<sup>8</sup> from fossilized tracks uncovered so far suggest that large bipedal dinosaurs moved at speeds of less than 5 m s<sup>-1</sup>. But it may be that tracks left by faster-moving dinosaurs just haven't been discovered yet.

In their analysis of *Tyrannosaurus*, Hutchinson and Garcia<sup>5</sup> introduce an approach based on estimates of the minimum muscle mass needed for fast running. First they applied their analysis to alligators and chickens — two living relatives of bipedal dinosaurs. The results show that alligators have less than half the muscle mass that they would need to run fast (if, like bipedal dinosaurs, they used only their hind limbs), whereas chickens have nearly twice the necessary hind-limb muscle mass. This agrees with the observed fact that chickens and many other avian bipeds are good runners, but alligators must support themselves on four limbs and move at relatively modest speeds.

Hutchinson and Garcia then extended their analysis to estimate the limb muscle mass of extinct animals and quantify their locomotive performance. From fossil specimens of *Tyrannosaurus*, the authors estimated body and segment mass, worked out areas of muscle attachment, and deduced the forces and moments that the creature's leg muscles could have generated. Their analysis rests on assumptions about the limb posture and the magnitude of reaction forces exerted by the ground on the limbs of *Tyrannosaurus*, and about the kinematic similarity between dinosaurs and living birds and mammals<sup>9</sup>. But their results show that, even if the creature used all its hind-limb muscle mass, it could not have generated the forces necessary for running. They show that for a chicken scaled up to 6,000 kg to run, it would need muscles in each leg equivalent to 99% of its body mass — which is obviously impossible. The results for smaller bipeds, however, show they probably could run quickly, in agreement with estimates of their speeds from fossil tracks<sup>8</sup>.

A pleasing aspect of Hutchinson and

Garcia's study is that they apply sensitivity analysis — allowing for a degree of parameter uncertainty — to evaluate the robustness of their results. For instance, they find that the estimated muscle mass is very sensitive to differences in limb posture, but is less sensitive to other parameters, such as muscle-fibre length. Collectively, these uncertainties contribute to up to a threefold variation in the estimated muscle mass for the various models of *Tyrannosaurus*, but the conclusion is unchanged. Palaeontological analysis of functional performance in fossil organisms will always be an uncertain science, dependent on the availability of fragmentary, long-dead material. So it is welcome when new analytical approaches such as this, and others (such as finite element analysis<sup>10</sup>), are brought to bear on such problems.

But what of the reputation that *Tyrannosaurus* has as a fearsome hunter? Hutchinson and Garcia's results suggest that the creature would have had little success chasing smaller, more fleet-footed prey; it may even have fed on carrion. But I suspect that it could still have moved fast enough to attack other large dinosaurs whose locomotive ability was also limited.

The dinosaurs are famous for being the largest creatures that ever inhabited our planet. But as a group they represent a broad range of size and diversity of form, with a similarly wide range of locomotive capacities and lifestyles<sup>11,12</sup>. It will be interesting to see what insights future investigations of dinosaur diversity yield as new analytical and computational approaches are explored. ■

Andrew A. Biewener is at the Concord Field Station, Harvard University, Old Causeway Road, Bedford, Massachusetts 01730, USA.

e-mail: abiewener@oeb.harvard.edu

1. Alexander, R. McN. *Dynamics of Dinosaurs and Other Extinct Giants* (Columbia Univ. Press, New York, 1989).
2. Farlow, J. O., Smith, M. B. & Robinson, J. M. *J. Vert. Paleontol.* **15**, 713–725 (1995).



**Figure 1 Was *Tyrannosaurus* as fleet of foot as we thought? Hutchinson and Garcia<sup>3</sup> analysed the muscle mass and forces in the legs of alligators and chickens, and then extrapolated their results to a 6,000-kg tyrannosaur. Their findings fly in the face of Hollywood legend — *Tyrannosaurus* did not have enough leg muscle to run.**

3. Bakker, R. T. *Dinosaur Heresies* (William Morrow, New York, 1986).
4. Paul, G. S. *Predatory Dinosaurs of the World* (Simon & Schuster, New York, 1988).
5. Hutchinson, J. R. & Garcia, M. *Nature* **415**, 1018–1021 (2002).
6. Biewener, A. A. *Science* **245**, 45–48 (1989).
7. Biewener, A. A. *Science* **250**, 1097–1103 (1990).
8. Farlow, J. O., Gatesy, S. M., Holtz, T. R., Hutchinson, J. R. & Robinson, J. M. *Am. Zool.* **40**, 640–663 (2000).
9. Alexander, R. McN. & Jayes, A. S. *J. Zool.* **201**, 135–152 (1983).
10. Rayfield, E. J. *et al. Nature* **409**, 1033–1037 (2001).
11. Carrano, M. T. *Paleobiol.* **24**, 450–469 (1998).
12. Middleton, K. M. & Gatesy, S. M. *Zool. J. Linn. Soc.* **128**, 149–187 (2000).

## Earth science

# Slip-sliding away

Steven N. Ward

**The side of an oceanic volcano, one of the Hawaiian islands, has been caught sliding towards the sea. The distance concerned was only a few centimetres. But it could be an indicator of a huge landslide to come.**

Seeing is believing. For Earth scientists especially, the adage holds considerable weight because the 'seeing' is so rare. The geological record tells us that mountain ranges have been built and then washed down to the sea; that entire ocean basins have opened and closed like a door; and that ice a mile thick blanketed the globe a dozen times over. Scientists believe that these events took place, but still, they

are hard to imagine. For most people such incidents might smack more of science fiction than science fact — but then, seeing is believing.

Who, having viewed film of Mount Pinatubo exploding, or the aftermath of a large earthquake, doesn't accept that 'big stuff' really does happen? In their paper on page 1014 of this issue<sup>1</sup>, Cervelli *et al.* provide a glimpse of what might end up to be big

## news and views

stuff — the whole side of an oceanic volcano falling into the sea, an event known as flank collapse.

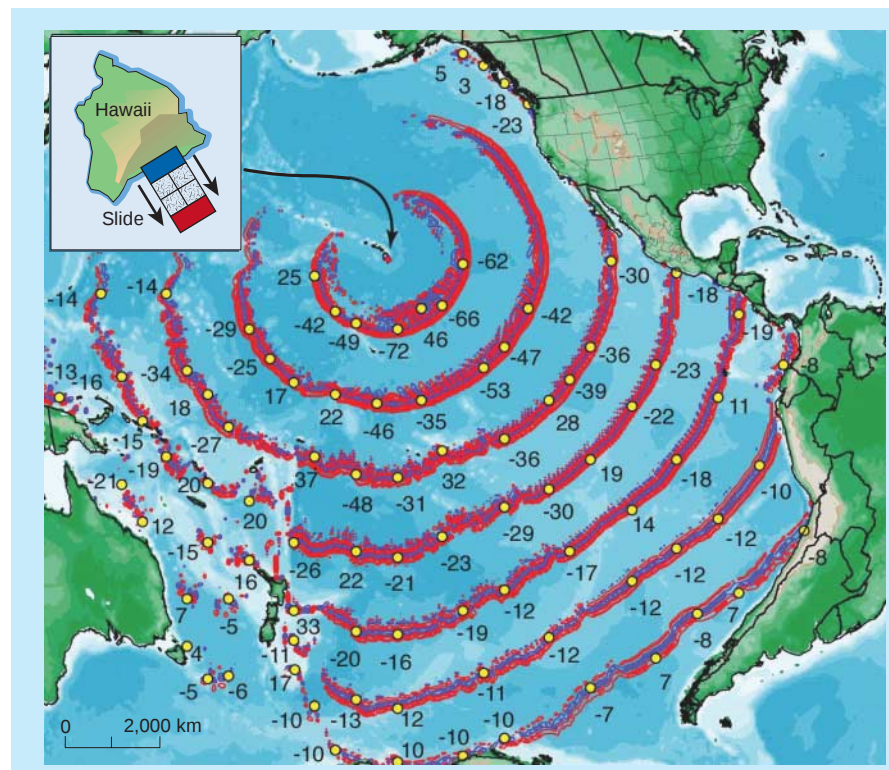
Over time, virtually all oceanic volcanoes grow, become too steep, and slough off flank material. We know this to be true because sonar surveys around most volcanic island chains reveal dozens of old, overlapping debris fields. The Hawaiian islands alone host 70 collapse fields dating from 20 million years ago. Adding up the number of debris fields from all of the ocean's volcanic islands yields the estimate that one flank collapse happens somewhere in the world every 10,000 years on average. Flank collapses are nature's great landslides. Embracing up to 5,000 km<sup>3</sup> of rock, they compare to a one-and-a-half-kilometre-thick slice of the state of Rhode Island or the island of Majorca racing sideways for 30 or 60 km. In contrast, when Mount Saint Helens erupted in 1980, only 3 km<sup>3</sup> of material blew away.

While flank collapses of oceanic volcanoes are common geologically, none has been caught in action — until now. Cervelli and colleagues' glimpse<sup>1</sup> of the action came from a network of 20 continuously recording stations of the global positioning system (GPS) scattered about the southeast slope of Kilauea volcano on Hawaii's big island. In November 2000, over a 36-hour period, these GPS stations witnessed a 20-km-long and 10-km-wide chunk of the southeast flank move seaward at the speed of 6 centimetres per day. For geophysicists accustomed to tectonic motions of a few millimetres per year, a few centimetres per day is like rocket travel.

To soothe any doubting Thomas, Cervelli *et al.* spend half of the report reviewing the details of their analysis. The effort is certainly thorough enough to dispel any notion that the signal is a fluke or masquerading noise. For me, their map of a dozen GPS displacement arrows (Fig. 1 on page 1015) all pointing out to sea far beyond their error ellipses tells the whole story. What else can they indicate but some early stage of one of those flank collapses that litter the geological record? A 2,000-km<sup>3</sup> piece of Hawaii is slip-sliding away.

In terms of predicting a collapse, the authors interpret their observations more cautiously than I do. By means of dislocation modelling, however, they confirm that the observed GPS displacement field could be explained by 10 cm of offset on a shallow dipping surface that lies 4.5 km under their network and that probably extends well out to sea. The offset was a silent earthquake, if you will, on the fault that may eventually detach the whole flank. Thankfully, the November 2000 slide stopped short, but what would it take to dislodge the whole block? Experts believe<sup>2–5</sup> that intense intrusion of the flank by molten magma dikes





**Figure 1** Computer simulation<sup>7</sup> of the tsunami waves that might be set off in a collapse of Kilauea's southeast flank. The simulation assumes that a block 40 km long, 20 km wide and 2,000 m thick (blue, inset) slides for 60 km at 70 m s<sup>-1</sup>. The tsunami wave fronts are pictured at two-hour intervals from 2 to 14 hours. Red and blue contours are wave elevations and depressions, respectively, and the numbers are sample wave heights in metres. Tsunamis from this collapse would have 1.72 10<sup>19</sup> J of energy (equivalent to about 4,100 megatons of TNT), and focus slightly towards the southeast. The waves span 280° of arc, sparing only locations to Hawaii's north and west. Tsunamis from volcanic flank collapses can dwarf those generated by earthquakes of any plausible magnitude. The model predicts potentially devastating 30-m waves beaching on the west coast of North America. These decay to 10 m in height by the time of their arrival at the tip of South America.

might provide the extra nudge, especially if the detachment fault is also lubricated by the injection of high-pressure groundwater.

Last year, Simon Day and I published an article<sup>6</sup> on the tsunami waves that might be generated by the collapse of another oceanic island volcano, Cumbre Vieja on La Palma in the Canary Islands. Our computer simulations predicted that a tsunami stirred by a 500-km<sup>3</sup> landslide there could span the entire Atlantic basin, keeping amplitudes of 10–20 metres. Based on these calculations, if a 2,000-km<sup>3</sup> piece of Kilauea ever does push into the sea, it could, under certain conditions, parent a tsunami that will strike much of the Pacific Rim (Fig. 1). Historical time has not seen a tsunami of this scale, but many researchers argue that geological deposits and landform shapes preserve the signature of older ones.

The implication that volcanic island collapses could raise extensive tsunamis grips both one's imagination and concern. Could potential collapses be monitored and perhaps forecast? Cervelli *et al.* demonstrate that current GPS technology deployed in networks at 5-km intervals can provide real-time detection of even seismically silent

shifts in volcanic edifices: small, silent shifts that may presage a fully fledged flank failure. The world's oceanic volcanoes are stages best not left unwatched. In a few years, dedicated radar satellites may take up volcano guard duty. For now, GPS provides one of the sharpest views.

People should not lose sleep over large but rare natural hazards. They should not run blind either, particularly when a useful eye exists. Until the next volcanic island does collapse we will never know how nature's great landslides play out, but for me, Cervelli and colleagues' article supports the case that seeing is believing.

Steven N. Ward is at the Institute of Geophysics and Planetary Physics, University of California at Santa Cruz, Santa Cruz, California 95064, USA.  
e-mail: ward@es.ucsc.edu

1. Cervelli, P., Segall, P., Johnson, K., Lisowski, M. & Miklius, A. *Nature* **415**, 1014–1018 (2002).
2. Elsworth, D. & Voight, B. *J. Geophys. Res.* **100**, 6005–6024 (1995).
3. Iverson, R. M. *J. Volcanol. Geotherm. Res.* **66**, 295–308 (1995).
4. Day, S. J. *J. Geol. Soc. Lond. Spec. Publ.* **110**, 77–93 (1996).
5. Elsworth, D. & Day, S. J. *J. Volcanol. Geotherm. Res.* **94**, 323–340 (1999).
6. Ward, S. N. & Day, S. *Geophys. Res. Lett.* **28**, 3397–3400 (2001).
7. Ward, S. N. *J. Geophys. Res.* **106**, B6, 11,201–11,216 (2001).

## Daedalus

### Support for neutrons

Nuclear matter, the stuff that neutron stars are made of, is essentially a mass of neutrons packed together. It weighs millions of tons per cubic millimetre; even a microscopic quantity would be hard to keep at the Earth's surface. Last week Daedalus argued that neutron-rich material would be stable, as neutron stars are. He is now working out how to hold it against gravity.

He planned to make it by persuading the electrons of an atom to dive into the nucleus. If not all of them did this, the result would still be normal atoms, with electrons orbiting the space outside each nucleus. Each atom would be superheavy, its dense nucleus stabilized by a vast excess of neutrons. The resulting solid would be a hundred times as dense as water, and very strong. The jewel in the crown of such materials, and the strongest of them all, would be superheavy diamond.

Daedalus recalls the Eiffel Tower, which has a sharply diverging base that supports its top load. He is designing a tiny 'Eiffel Pyramid', topped with a microscopic sliver of pure nuclear matter. The wider layers beneath are dense, strong, intermediate matter; under these are wider layers of less dense matter; the broad base of the Pyramid is less dense still, the strongest normal matter. The Eiffel Pyramid could test Daedalus's bold prediction that nuclear matter should absorb neutrinos.

Normal matter is mainly empty space, so neutrinos go straight through it; indeed, billions of neutrinos a second penetrate all normal matter in every direction. Even the Sun, from which neutrinos escape so easily, has such a low density that it would float in strong potassium carbonate solution. But nuclear matter is over 10<sup>15</sup> times denser. It should absorb the neutrino background. Indeed, black holes and neutron stars could be the only neutrino sinks known.

Daedalus would love to exploit our own neutrino background in this way. He will be alert for warming at the top of his Eiffel Pyramid of nuclear matter — a sign that neutrinos are being absorbed. How splendid it would be to capture neutrinos, the ultimate waste product of the Universe, and use their energy for our own needs. The neutrino flux would be useful even as background heating; if it generates temperatures above 100 °C it could be used to raise steam and power. Compared with the vast sums expended on the magnetic fusion reactor ITER, and the ever-receding dream of fusion power, the idea of exploiting free solar or cosmic neutrinos looks quite attractive.

David Jones

# Mellifluous matures to malodorous in musth

Mood-altering secretions by excited male elephants smooth out social interactions.

**M**ale Asian elephants in musth — an annual period of heightened sexual activity and intensified aggression — broadcast odoriferous, behaviourally influential messages from secretions of the temporal gland<sup>1–4</sup>. From our observations in the wild, together with instantaneous chemical sampling and captive-elephant playback experiments, we have discovered that young, socially immature males in musth signal their naivety by releasing honey-like odours to avoid conflict with adult males, whereas older musth males broadcast malodorous combinations to deter young males, facilitating the smooth functioning of male society. As elephant–human conflicts can upset this equilibrium, chemically modulating male behaviour may be one way to help the conservation of wild elephants.

During musth in adult male Asian elephants (*Elephas maximus*; Fig. 1), serum pH is raised and androgen synthesis, ketogenesis and lipid catabolism all increase<sup>2–4</sup>. These physiological alterations are reflected in malodorous secretions that stream from the elephant's unique temporal gland<sup>4</sup>.

Ancient Hindu poetry describes the arrival of bees to “gather sweetness from the temples of [young] musth elephants”<sup>5</sup>.

In keeping with this poetic observation, we were struck by the powerful honey-like odour emitted by a captive 11-year-old male Asian elephant when he began secreting. This youthful musth was termed ‘moda’, a period that is characterized by mischievous and erratic behaviour<sup>6</sup>. We therefore investigated whether the extensive dual olfactory systems of male Asian elephants can differentiate between youthful musth chemical signals and those from sexually and socially mature males.

We analysed the changing chemical composition of temporal-gland secretions from young post-pubertal male Asian elephants, who were exhibiting episodes of aberrant behaviour and had fluctuating hormone levels (Table 1). Combining innovative modification of solid-phase micro-extraction<sup>7</sup>, capture of headspace volatiles in evacuated bottles<sup>8</sup>, and chemical analysis by gas chromatography with mass spectrometry<sup>7,8</sup>, we identified the predominant components in the mellifluous exudates as 3-hexen-1-ol, two ketones (2-heptanone and acetophenone) and various acetates (mainly isoamyl acetate) — all of which are sweet-smelling (Table 1). There is a marked chemical convergence between the temporal-gland secretions of male Asian elephants in



**Figure 1** A mature musth male takes a swim to cool off. During musth, secretions emanate from the temporal gland near the eye.

L. E. L. RASMUSSEN

moda musth, and honeybees and honey — honey also contains 3-hexen-1-ol and sweet-smelling acetates, and bees use pheromone blends containing 2-heptanone and acetates, including isoamyl acetate<sup>9</sup>.

As males mature (post-moda), their periods of increased androgen hormone levels lengthen, aggressive behaviour becomes overt, and their copious secretions are chemically altered to resemble those of the adult temporal gland<sup>3</sup> (Table 1). Concentrations of 3-hexen-1-ol and acetates are greatly reduced, and acetophenone is replaced by a foul-smelling combination of 2-nonanone, acyclic ketones and substituted cyclohexanones, including 3-methyl-2-cyclohexen-1-one and frontalin (1,5-dimethyl-6,8-dioxabicyclo[3.2.1]octane)<sup>4,8</sup>. These last two compounds are also secreted together as pheromones by bark beetles (*Dendroctonus tenebrans*)<sup>10,11</sup>.

‘Playback’ experiments monitoring the responses of captive males to collected samples of temporal-gland secretions, together with observations of the behaviour of wild males of known chemical status, supported our idea that chemical signals change in association with maturing musth-linked behaviours. Moda males generally avoided the secretions of older males, whereas samples from moda males elicited little response from mature individuals (Table 1, and see supplementary information).

Encounters observed in the wild between a focal musth male, emitting chemically characterized adult secretions, and 12 different male conspecifics, emitting varying chemical signals, were consistent with our results from elephants in captivity. After sniffing the focal musth male from distances of 3–100 metres, three juveniles showed little reaction, five young males acquiesced, and four were apprehensive, choosing to avoid the senior male.

When the smell of a chemically defined moda male was assessed from a distance by

**Table 1** Temporal-gland secretion signalling by elephants in musth

| Male elephants                           | Honey moda                         | Post-moda                      | Older adult musth                                             |
|------------------------------------------|------------------------------------|--------------------------------|---------------------------------------------------------------|
| Age (years)                              | 8–13                               | 14–18                          | 25–35                                                         |
| Behavioural status†‡                     | 1.0 (1.0–4.0)* <sup>1,2</sup>      | 17.5 (10.0–25.0)* <sup>2</sup> | 34 (32.0–39.0)* <sup>1</sup>                                  |
| Serum androgens (ng ml <sup>-1</sup> )†§ | 13.2 (10.3–16.8)* <sup>1</sup>     | 21.9 (11.0–30.0)* <sup>2</sup> | 36.7 (31.9–65.6)* <sup>1,2</sup>                              |
| n, NS                                    | 4, 32                              | 4, 14                          | 6, 18                                                         |
| <b>TGS compounds</b>                     |                                    |                                |                                                               |
|                                          | Concentration (g m <sup>-3</sup> ) |                                |                                                               |
| Esters†                                  | 10.5 (8.5–13.5)¶                   | ND                             | ND                                                            |
| 3-Hexen-1-ol†                            | 30.0 (8.5–13.5)                    | ND                             | ND                                                            |
| Acetophenone†#                           | 2.0 (1.05–2.15)* <sup>1,2</sup>    | 0.15 (0.1–0.2)* <sup>2</sup>   | 0.095 (0.06–0.10)* <sup>1</sup>                               |
| Frontalin†‡                              | ND                                 | 0.10 (0.10–0.12)*              | 0.22 (0.20–0.30)*                                             |
| 2-Nonanone†\$                            | ND                                 | 0.10 (0.05–0.10)*              | 0.23 (0.21–0.30)*                                             |
| <b>State of responder</b>                |                                    |                                |                                                               |
|                                          | TGS playbacks                      |                                |                                                               |
| Moda (n = 3)                             | 89% attracted<br>(3 tests)         | NT                             | 80% repelled<br>(5 tests)                                     |
| Adult (n = 3)                            | 78% no response<br>(3 tests)       | NT                             | 33% attracted<br>33% no response<br>33% repelled<br>(5 tests) |

Behavioural and chemical evidence for signalling using temporal-gland secretion (TGS) by male Asian elephants (median, 25–75%) during moda and adult musth states (see supplementary information). n, number of elephants; ND, not detected; NS, number of samples; NT, not tested. Asterisks represent significant results.

†Median (25–75%).

‡Scored 0 (no deviation from normal) to 5 (maximum deviation from normal); P < 0.05. \*DR (difference of ranks) = 37.9, Q = 6.99;

\*<sup>2</sup>DR = 23.9, Q = 4.05.

\$P < 0.0001, d.f. = 2, H = 28.8. \*<sup>1</sup>DR = 29.33, Q = 5.35;

\*<sup>2</sup>DR = 21.60, Q = 3.26.

¶Number of esters (see supplementary information).

||Detectable limit is > 1 part per billion volume.

#P < 0.0001, d.f. = 2, H = 48.9; \*<sup>1</sup>DR = 35.37, Q = 6.46;

\*<sup>2</sup>DR = 27.65, Q = 4.66.

‡U = 74, P < 0.003 (see supplementary information).

\$U = 74, P < 0.0001.

Kruskal-Wallis ANOVA, Dunn's pairwise multiple comparison (§\$#);

Mann-Whitney U-test (‡‡\$).



the focal male, it was ignored, suggesting that the honey odour conveys a non-threatening message. However, the focal male stalked and attacked a different, slightly secreting young adult in a dominance interaction (see supplementary information). It is evidently advantageous to be able to recognize the ontogenic degree of musth in conspecifics before initiating physical encounters<sup>3,12,13</sup>.

Unravelling this medley of chemical signals helps to clarify the behavioural and physiological mechanisms that underlie the phenomenon of musth and its influence on other males. This knowledge should help in the formulation of deterrence programmes in southern India against crop-raiding wild elephants, most of which are male and are often in musth. Moreover, the moda-musth emanations of young maturing elephants, as poetically observed by the ancient Hindus, have now been substantiated by modern scientific techniques.

**L. E. L. Rasmussen\*, H. S. Riddle†, V. Krishnamurthy‡**

\*Department of Biochemistry and Molecular Biology, OGI School of Science and Engineering,

Oregon Health and Science University, Beaverton, Oregon 97006-8921, USA

e-mail: betsy@bmb.ogi.edu

†Riddle's Elephant and Wildlife Sanctuary,

PO Box 715, Greenbrier, Arkansas 72058, USA

‡The Asian Elephant Research and Conservation Centre, Indian Institute of Science, Bangalore 560 012, India

1. Jainudeen, M. R., McKay, G. M. & Eisenberg, J. G. *Mammalia* **36**, 247–261 (1972).
2. McKay, G. M. *Smithson. Contr. Zool.* **125**, 1–113 (1973).
3. Rasmussen, L. E. L. & Perrin, T. E. *Physiol. Behav.* **67**, 539–549 (1999).
4. Rasmussen, L. E. L. *Ecoscience* **5**, 410–426 (1998).
5. Edgerton, F. *The Elephant-Lore of the Hindus* (Motilal Banarsidass, Delhi, 1931).
6. Chandrasekharan, K. et al. in *The Asian Elephant: Ecology, Biology, Disease, Conservation and Management* (eds Silas, E., Nair, M. & Nirmalan, G.) 71–74 (Kerala Agric. Univ., India, 1992).
7. Arthur, C. L. & Pawliszyn, J. *Anal. Chem.* **62**, 2145–2148 (1990).
8. Perrin, T. E., Rasmussen, L. E. L., Gunawardena, R. & Rasmussen, R. A. *J. Chem. Ecol.* **21**, 207–221 (1996).
9. Free, J. B. *Pheromones of Social Bees* (Chapman & Hall, London, 1987).
10. Francke, W. et al. *J. Chem. Ecol.* **21**, 1043–1064 (1995).
11. Phillips, T. W., Nation, J. L., Wilkinson, R. C. & Foltz, J. L. *J. Chem. Ecol.* **15**, 1513–1533 (1989).
12. Krishnan, M. *J. Bombay Nat. Hist. Soc.* **69**, 298–315 (1972).
13. Sukumar, R. *The Asian Elephant: Ecology and Management* (Cambridge Univ. Press, Cambridge, 1989).

Supplementary information accompanies this communication on Nature's website (www.nature.com).

†United Institute of Geology, Geophysics and Mineralogy, Siberian Branch of the Russian

Academy of Sciences, Novosibirsk 630090, Russia

‡Institute of Geochemistry, Siberian Branch of the Russian Academy of Sciences, Irkutsk 664033, Russia

1. BDP Members *Russian Geol. Geophys.* **41**, 3–32 (2000).
2. Antipin, V. et al. *Quat. Int.* **80–81**, 19–36 (2001).
3. Williams, D. F. et al. *Science* **278**, 1114–1117 (1997).
4. Kashiwaya, K., Ochial, S., Sakai, H. & Kawai, T. *Nature* **410**, 71–74 (2001).
5. Moore, T. C., Klitgord, K. D., Golmstok, A. J. & Weber, E. *GSA Bull.* **109**, 746–766 (1997).

**Kashiwaya et al. reply** — The most important issue in the dating of the long Baikal core (600 m) is whether the segment between the base of C3An.2n and the top of C3Bn (267.67–375.48 m) is distorted. Changes in  $\gamma$ -ray intensity (H. Tsukahara et al., personal communication), which reflects the structure of the cores, indicate that this part is different from the rest (strictly speaking, the upper and lower shifting points seem to be at around 262 m and 362 m, respectively, from the fluctuation). To investigate this difference, we carried out spectral analyses of the upper section of the core (163–261 m), and of the middle (263–361 m) and lower (363–673 m) parts. Prevailing periods for each part are different, particularly in the middle one. There is a distinct common prevailing period of around 4.5–4.9 m in the upper and lower parts, whereas a period of 18 m prevails in the middle part, suggesting that the structure of the middle is different from the upper and lower parts of the core.

We are therefore reluctant to propose an age model that includes the middle part without further information on this section. Another recent age model<sup>1</sup> omits any discussion of this point, although the structure cannot be explained without further information.

**Kenji Kashiwaya\*, Shinya Ochial\*, Hideo Sakai†, Takayoshi Kawai‡**

\*Department of Earth Sciences, Kanazawa University, Kakuma, Kanazawa 920-1192, Japan

e-mail: kashi@kenroku.kanazawa-u.ac.jp

†Department of Earth Sciences, Toyama University, Toyama 930-8555, Japan

‡National Institute for Environmental Studies, Tsukuba, Ibaragi 350-0044, Japan

1. Antipin, V. et al. *Quat. Int.* **80–81**, 19–36 (2001).

## COMMUNICATIONS ARISING

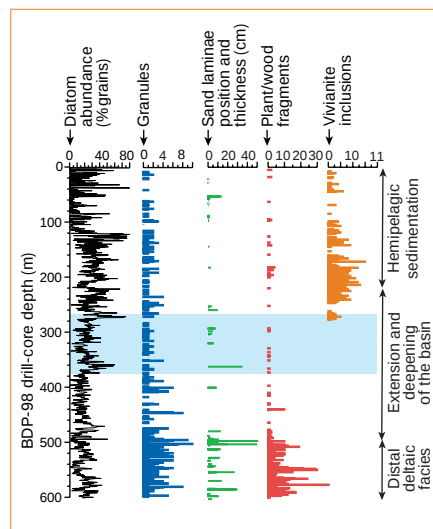
### Geomorphology

## Age of long sediment cores from Lake Baikal

**T**he new BDP-98 drill core of the Baikal Drilling Project is a key palaeoclimate record in continental Asia because globally sensitive sedimentary records of such length and continuity are very rare<sup>1–3</sup>. Kashiwaya et al.<sup>4</sup> have attempted signal processing of the BDP-98 average grain-size record, but in constructing their age model they excised a 100-metre interval from the 600-metre section, stating that it is “erroneous”<sup>4</sup>. On the basis of our lithological studies, we consider that this excision is unjustified.

Moreover, the interval excised by Kashiwaya et al.<sup>4</sup> corresponds to the spatially continuous sedimentary unit between seismic reflection boundaries B7 and B8 at Academician Ridge, Lake Baikal<sup>2,5</sup>. In addition, the lithology of the BDP-98 section reflects a progressive change of facies from shallow-water pro-deltaic to deep-water hemipelagic sedimentation at the drill site (Fig. 1). These findings imply that the sedimentation rates in the lower part of the BDP-98 are significantly higher than those assumed by Kashiwaya et al.<sup>4</sup> and that the bottom age of the BDP-98 is significantly younger, perhaps around 9–10 Myr (ref. 2).

We agree that the long-term insolation cycles are imprinted in the Lake Baikal palaeoclimate proxy records<sup>3</sup>, but our litho-



**Figure 1** Progressive change of facies at the BDP-98 drill site<sup>2</sup>, indicating that average sedimentation rates cannot be almost uniform throughout the section, as suggested by the age model of Kashiwaya et al.<sup>4</sup>. The lithology of the section also shows that there is no justification for the excision of the shaded interval in order to fit the age model. All of the components, apart from diatom abundance and sand laminae, are expressed as occurrences per 1-m interval in the split core<sup>3</sup>.

logical results indicate that there is a need to revise the age model and the signal-processing results of Kashiwaya et al.<sup>4</sup>.

**Alexander A. Prokopenko\*†, Eugene B. Karabanov\*‡, Douglas F. Williams\***

\*Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

e-mail: sasha@geol.sc.edu

**brief communications** is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from [nature@nature.com](mailto:nature@nature.com)

# MAP kinase signalling cascade in *Arabidopsis* innate immunity

Tsuneaki Asai<sup>\*§</sup>, Guillaume Tena<sup>\*§</sup>, Joulia Plotnikova<sup>\*</sup>, Matthew R. Willmann<sup>\*</sup>, Wan-Ling Chiu<sup>\*</sup>, Lourdes Gomez-Gomez<sup>†</sup>, Thomas Boller<sup>‡</sup>, Frederick M. Ausubel<sup>\*</sup> & Jen Sheen<sup>\*</sup>

<sup>\*</sup> Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

<sup>†</sup> Instituto de Desarrollo Regional, Sección de Biotecnología, Campus Universitario s/n, E-02071 Albacete, Spain

<sup>‡</sup> Friedrich Miescher-Institute, PO Box 2543, CH-4002 Basel, Switzerland

<sup>§</sup> These authors contributed equally to the work

**There is remarkable conservation in the recognition of pathogen-associated molecular patterns (PAMPs) by innate immune responses of plants, insects and mammals. We developed an *Arabidopsis thaliana* leaf cell system based on the induction of early-defence gene transcription by flagellin, a highly conserved component of bacterial flagella that functions as a PAMP in plants and mammals. Here we identify a complete plant MAP kinase cascade (MEKK1, MKK4/MKK5 and MPK3/MPK6) and WRKY22/WRKY29 transcription factors that function downstream of the flagellin receptor FLS2, a leucine-rich-repeat (LRR) receptor kinase. Activation of this MAPK cascade confers resistance to both bacterial and fungal pathogens, suggesting that signalling events initiated by diverse pathogens converge into a conserved MAPK cascade.**

The innate immune system is the first line of inducible defence against infectious disease. A key function of innate immunity is the detection of PAMPs produced by infectious agents but not by host cells<sup>1–3</sup>. In insects and mammals, the recognition of PAMPs is often mediated by Toll and Toll-like receptors (TLRs), respectively, with extracellular LRRs<sup>2,4</sup>. In plants, LRR domains are also found in the products of disease resistance genes<sup>5–8</sup>, which are receptors important for innate immunity.

Recent studies have identified bacterial flagellin as a PAMP that is recognized by the innate immune system in diverse organisms including insects, mammals and plants<sup>9–11</sup>. In mammals, TLR5 mediates the innate immune responses to flagellin<sup>11</sup>. In *Arabidopsis*, the LRR receptor kinase FLS2 is required for flagellin signalling<sup>12</sup>. A 22-amino-acid peptide flg22, corresponding to the most conserved domain of eubacterial flagellin, binds to FLS2 and induces defence responses in leaves<sup>10,12,13</sup>.

In contrast to animals, the molecular mechanisms underlying the responses to PAMPs are largely unknown in plants<sup>5–7</sup>. In mammals, different TLRs specifically recognize distinct PAMPs but activate common and conserved signal transduction pathways<sup>2,4</sup>. In plants, many different types of pathogens and pathogen-derived elicitors also trigger similar defence responses<sup>7,14–17</sup>. The best candidates for components of convergent signalling pathways in plants are the mitogen-activated protein kinases (MAPKs)<sup>7,16–19</sup>. Although several MAPKs involved in plant defence response have been identified<sup>14,20–25</sup>, the identity of the upstream receptors, MAPK kinases (MAPKKs) and MAPKK kinases (MAPKKKs), and the downstream transcription factors remain mostly unclear.

To elucidate MAPK signalling cascades in plant innate immune responses, we developed an *Arabidopsis* protoplast transient expression system in which transcription of early defence genes is activated by flg22 and in which the role of MAPK cascade components could be systematically tested and assigned. Using this system, we have identified a complete plant MAPK cascade and WRKY transcription factors acting downstream of the flagellin receptor FLS2. Our data also suggest that this signalling pathway functions in response to both fungal and bacterial pathogens and potentially could be engineered to enhance the disease resistance of crop plants to a wide range of pathogens.

## Early defence gene transcription

To dissect the early signal transduction pathways in plant innate

immune responses, we first established a leaf cell assay based on flg22 inducible transcription of early response genes in *Arabidopsis* mesophyll protoplasts<sup>26</sup>. Few reporter genes have been developed for the dissection of the early stages of defence signalling pathways in plants<sup>27–29</sup>. To identify genes that are induced by flg22-activated defence signalling, we created a subtracted complementary DNA library that represented messenger RNA species induced at various times by the elicitor flg22 in *Arabidopsis* mesophyll protoplasts<sup>30</sup>. Using flg22 rather than a pathogen or natural elicitors avoided the possibility that multiple elicitors could be functioning in parallel<sup>12</sup>. Furthermore, synchronous elicitation by flg22 is achieved more reproducibly in homogeneous mesophyll protoplasts than in intact leaves<sup>26</sup>. In the library we found well described defence genes, such as *PAL1* (At2g37040 in the *Arabidopsis* Genome Initiative nomenclature), *GST1* (Atlg02930), *PR1* (At2g19990) and *PR5* (Atlg75040), which are induced by a variety of pathogens and elicitors at different stages of the defence response in many plant species<sup>28,29,31</sup>. These genes were also induced in *Arabidopsis* leaves infiltrated with flg22 (not shown), suggesting that similar defence responses are induced by flg22 in isolated leaf protoplasts and in leaves of intact plants.

To develop new reporter genes that are expressed early in the primary defence response, we focused on flg22-activated genes identified in the subtraction library that are expressed at early time points (see below) and code for putative regulatory factors, including *WRKY29* (At4g23550)<sup>32</sup> and *FRK1* (*FLG22-INDUCED RECEPTOR-LIKE KINASE 1*; At2g19190), which encode a WRKY transcription factor (with a conserved WRKY DNA-binding domain) and a LRR receptor kinase, respectively. Induction of WRKY transcription factors and LRR receptor kinase genes by pathogens, pathogen-derived elicitors or salicylic acid has been demonstrated in the leaves of several plant species including *Arabidopsis*<sup>28,29,32–34</sup>. These results indicate that *Arabidopsis* protoplasts treated with flg22 increase transcription of defence-related genes that are also expressed in intact plants after pathogen infection.

Polymerase chain reaction after reverse transcription of RNA (RT-PCR) analysis showed that *WRKY29*, *FRK1* and *GST1* mRNA levels were increased in *Arabidopsis* protoplasts within 30 min after flg22 treatment (Fig. 1a), whereas induction of the extensively studied *PR1* and *PR5* genes occurred much later (Fig. 1a). Two stress-regulated genes, H<sub>2</sub>O<sub>2</sub>-inducible *GST6* (At2g47730) and ABA-, cold-, or drought-responsive *RD29A* (AT5g52310)<sup>35</sup>, were



not induced by flg22 over basal levels of expression (Fig. 1a). In contrast to the common assumption that protoplasts isolated from fresh leaves are badly damaged, stressed and dying, or are already activated in defence responses, these data show that *Arabidopsis* mesophyll protoplasts respond specifically to flg22, similarly to the response observed in intact plants<sup>13</sup>.

To examine whether *de novo* protein synthesis is required for flg22-induced defence gene expression, protoplasts were preincubated with the protein synthesis inhibitor cycloheximide (CHX) and then treated with flg22. RT-PCR analysis showed that induction of the early response genes *WRKY29*, *FRK1* and *GST1* by flg22 was not significantly affected by CHX, whereas the inhibitor blocked induction of the late response genes *PR1* and *PR5*, but not the expression of other control genes (Fig. 1a). To determine whether the early-defence genes are activated transcriptionally, the promoters of the *WRKY29*, *FRK1*, *GST1*, *GST6* and *RD29A* genes were

fused to the firefly luciferase reporter gene (*LUC*)<sup>35</sup> and tested for their response to flg22 in transiently transfected protoplasts<sup>35</sup>. Consistent with the RT-PCR data obtained with the endogenous genes, the *WRKY29*, *FRK1* and *GST1* promoters, but not the *GST6* and *RD29A* promoters, were induced by flg22 (Fig. 1b). As a parallel control, the 23-amino-acid peptide flg23RM, corresponding to the flg22-homologous sequence of *Rhizobium meliloti* flagellin, activated none of these promoters (not shown). *Rhizobium meliloti* is a plant symbiont that fails to activate host innate immunity<sup>10,13</sup>. These results suggest that in response to pathogen-derived flagellin, *WRKY29*, *FRK1* and *GST1* are rapidly induced by pre-existing signalling molecules. Because the H<sub>2</sub>O<sub>2</sub>-responsive gene *GST6* (ref. 35) was not induced by flg22 (Fig. 1a, b), it is likely that although flg22 induces an oxidative burst in *Arabidopsis* leaves<sup>12,13</sup>, it may not elicit a strong or persistent oxidative burst in protoplasts. Thus, the flg22-protoplast system distinguishes between flagellin and oxidative-stress signalling pathways (see below). In addition to the well established parsley protoplast system derived from suspension cultures<sup>7,15,16,20,33</sup>, the *Arabidopsis* mesophyll protoplast transient expression system offers a new tool to study defence signalling based on early gene transcription.

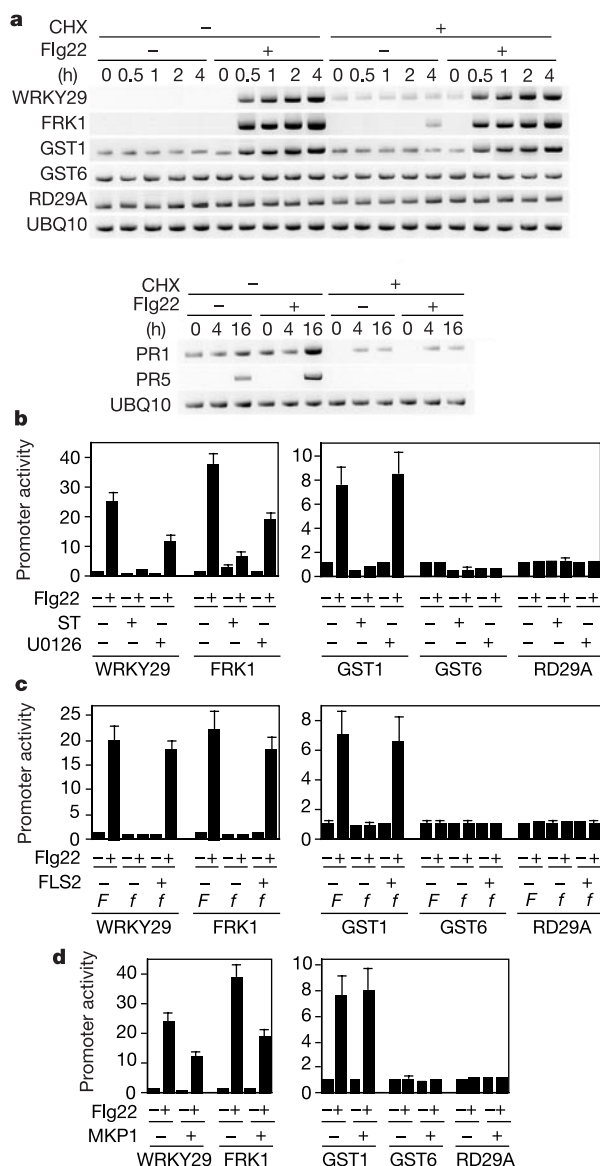
### Flg22 signals through FLS2

To determine whether the *WRKY29*, *FRK1* and *GST1* promoters were activated by flg22 through the FLS2 LRR receptor kinase<sup>12</sup>, protoplasts isolated from *fls2* mutant leaves were transiently transfected with the *WRKY29-LUC*, *FRK1-LUC* and *GST1-LUC* reporter constructs. In the *fls2* mutant protoplasts, none of the three promoters could be activated by flg22 (Fig. 1c). We note that transient expression of wild-type FLS2 (ref. 12) restored flg22-inducibility of the promoters in the *fls2* mutant protoplasts (Fig. 1c). Consistent with these results, the broad-spectrum protein kinase inhibitor staurosporine effectively blocked the ability of flg22 to activate the *WRKY29*, *FRK1* and *GST1* promoters (Fig. 1b). These data show that flagellin signalling leading to the expression of *WRKY29*, *FRK1* and *GST1* requires FLS2.

### Specific MAPKs in flg22 signalling

The role of MAPK signalling in the flg22-mediated activation of the *WRKY29*, *FRK1* and *GST1* promoters was tested by using the MAPKK inhibitor U0126. As shown in Fig. 1b, U0126 partially blocked transcription of the *WRKY29* and *FRK1* promoters but had no effect on the *GST1* promoter. To further test the involvement of MAPK signalling in the activation of *WRKY29* and *FRK1* promoters, a mouse MAPK phosphatase (MKP1) was transiently expressed in *Arabidopsis* protoplasts<sup>36</sup>. Similarly to U0126, MKP1 partly reduced flg22-induced activation of the *WRKY29* and *FRK1* promoters but had no effect on the *GST1* promoter (Fig. 1d). These results suggested that both MAPK-dependent and MAPK-independent signalling pathways act downstream of FLS2 to activate the *WRKY29* and *FRK1* promoters. Further investigation will be required to test whether calcium signalling plays a role in the MAPK-independent pathway induced by flg22, as suggested by the studies in parsley protoplasts<sup>15</sup>.

Because flg22 activation of the *WRKY29* and *FRK1* promoters was partially blocked by U0126 and MKP1, we examined whether flg22 can activate MAPKs in *Arabidopsis* protoplasts. Treatment of protoplasts with flg22 resulted in rapid activation of endogenous protein kinases that phosphorylated myelin basic protein (MBP), a commonly used MAPK substrate, in an in-gel kinase assay (Fig. 2a). These protein kinases were not activated by treatment with flg23RM (not shown), suggesting their importance in pathogen defence. To systematically test the *Arabidopsis* MAPKs (MPKs) to determine which one(s) are involved in flg22 signalling, we chose six representative MPKs corresponding to four out of the five MPK subfamilies that may exhibit distinct functions on the basis of sequence homology analysis<sup>9,37</sup>. These MPKs were tagged with



**Figure 1** Early-defence gene activation by flg22. **a**, RT-PCR analysis. Protoplasts were preincubated for 1 h without or with 10  $\mu$ M cycloheximide (CHX), and then treated without or with 100 nM flg22. **b**, Involvement of MAPK. Transfected protoplasts were preincubated for 1 h with 1  $\mu$ M staurosporine (ST) or with 1  $\mu$ M U0126 before induction by 100 nM flg22 for 4 h. **c**, FLS2 requirement. Protoplasts were isolated from wild-type (*F*) and *fls2-22* mutant (*f*) leaves. Transfected protoplasts were incubated for 9 h first before incubation without or with 100 nM flg22 for 4 h. **d**, MKP1 suppression.

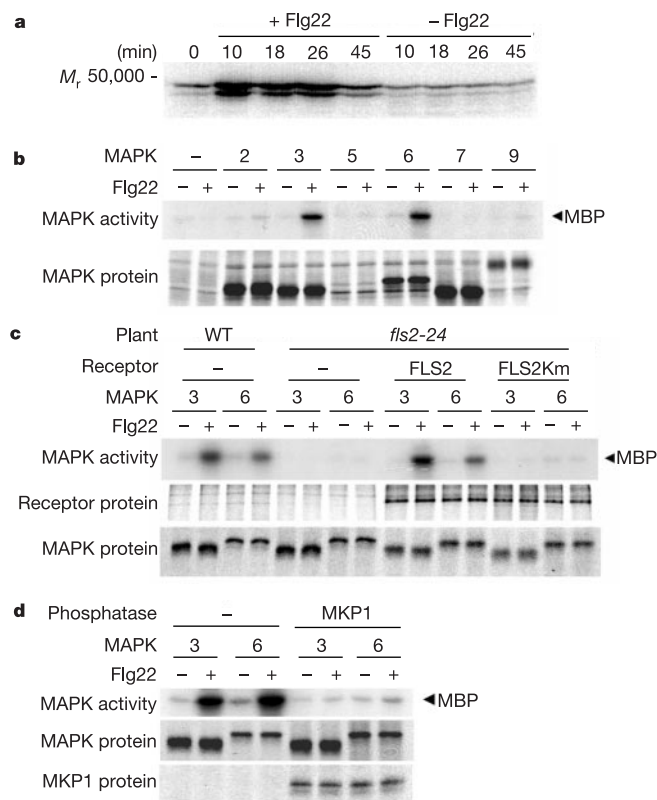
the haemagglutinin (HA) epitope, transiently expressed in protoplasts, immunoprecipitated with an anti-HA antibody, and tested *in vitro* for MBP kinase activity<sup>35</sup>. The results indicated that MPK3 and MPK6, which belong to the same subfamily, but not the other MPKs, showed strong activation following flg22 treatment (Fig. 2b). The involvement of MPK5 cannot be ruled out, however, owing to its poor expression and/or instability. The result is consistent with the previously reported activation of MPK6 in *Arabidopsis* cultured cells and leaf strips by flg22 (ref. 23). The activation of MPK3 and MPK6 by flg22 did not occur in the *fls2* mutant protoplasts unless functional FLS2, but not a kinase-inactive mutant of FLS2 (FLS2Km), was co-expressed (Fig. 2c). The co-expression of MKP1, which partially blocked flg22-induced activation of WRKY29 and FRK1 promoters (Fig. 1d), eliminated MPK3 and MPK6 activation by flg22 (Fig. 2d). These results suggest that flagellin perception by FLS2 leads to the activation of MPK3 and MPK6. The studies also revealed the importance of MAPK-independent pathway in flg22 signalling.

### Redundant MAPKKs in flg22 signalling

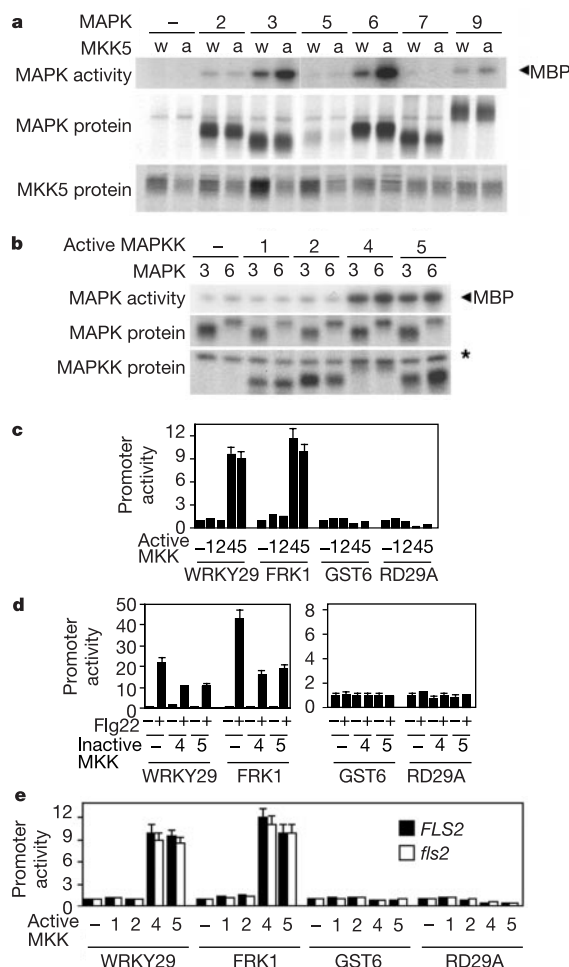
The *Arabidopsis* genome contains nine MAPKK (MKK) genes that belong to four subfamilies, suggesting at least four distinct functions<sup>19</sup>. Four of these MKKs, MKK1, MKK2, MKK4 and MKK5, belonging to two subfamilies, are expressed in leaf cells<sup>19,37</sup>. Therefore, these four MKK genes were cloned and analysed in the protoplast transient expression assay. To distinguish MKKs from the HA-tagged MPKs in transfected protoplasts, a Myc-

epitope tag was fused to each MKK. MAPKKs require phosphorylation to be activated, so gain-of-function mutants of the four MKKs were generated by converting the conserved serine and/or threonine residues in the kinase-activation loop located between subdomains VII and VIII to aspartate or glutamate to mimic phosphorylation. The Myc-tagged wild-type or constitutively active MKKs were co-expressed individually with the six HA-tagged MPKs to systematically determine their regulatory relationships. The results showed that constitutively active MKK4 and MKK5 (MKK4a and MKK5a, respectively) were equally effective at activating MPK3 and MPK6 (Fig. 3a, b). In contrast, constitutively active MKK1 and MKK2 (MKK1a and MKK2a, respectively) were unable to activate these MPKs (Fig. 3b), although MKK1a did activate other MPKs (not shown).

To test directly the role of MKK4 and MKK5 in flg22 signalling, plasmid DNA expressing the constitutively active MKK constructs was cotransfected with various reporter gene constructs into protoplasts<sup>35</sup>. Either MKK4a or MKK5a, but not MKK1a or



**Figure 2** Flg22 activates MPK3 and MPK6 through FLS2. **a**, Flg22 activates endogenous MBP kinases. Protoplasts were treated with or without 1  $\mu$ M flg22. **b**, Flg22 activates MPKs. MAPK activation with or without 100 nM flg22 for 10 min (top) and expression of each MAPK are shown (bottom). **c**, FLS2 requirement. MPK3 or MPK6 was co-expressed with a wild-type (FLS2) or kinase-inactive (FLS2Km) flg22 receptor in wild-type (WT) and *fls2-24* protoplasts. MAPK activities (top) and FLS2 (middle) and MAPK (bottom) expression are shown. **d**, MKP1 abolishes flg22 activation of MPKs. MAPK activity (top), and MAPK (middle) and MKP1 (bottom) expression are shown.



**Figure 3** MKK4 and MKK5 activate MPK3/MPK6 and early-defence genes. **a**, MKK5 activates MPK3 and MPK6. Protoplasts were expressing MAPK and wild-type (w) or constitutively active MKK5 mutant (a). MAPK activity (top) and MAPK (middle) and MAPKK (bottom) expression are shown. **b**, MKK4 and MKK5 redundancy. Protoplasts were expressing MPK3 or MPK6 and one constitutively active MKK. MAPK activity (top) and MAPK (middle) and MAPKK (bottom) expression are shown. Asterisk indicates a background band. MBP, myelin basic protein. **c**, MKK4a and MKK5a activate WRKY29 and FRK1 promoters. Protoplasts were expressing GFP or a constitutively active MKK. **d**, Dominant-negative MKK4 and MKK5 partially inhibit early-defence gene promoters. Protoplasts were expressing GFP or a kinase-inactive MKK4 or MKK5. **e**, MKK4 and MKK5 act downstream of FLS2. Wild-type (FLS2) and *fls2-24* (*fls2*) protoplasts were expressing GFP or a constitutively active MKK.



MKK2a, was sufficient to activate specifically the *WRKY29* and *FRK1* promoters (Fig. 3c). To test further the importance of MKK4 and MKK5 in the early-defence response, corresponding dominant-negative mutants were generated by replacing the conserved lysine residue in the ATP-binding site with methionine. As shown in Fig. 3d, dominant-negative MKK4 or MKK5 but not MKK1 (not shown) could partially block flg22 activation of the *WRKY29* and *FRK1* promoters similarly to U0126 or MKP1 (Fig. 1b, d). In addition, MKK4a and MKK5a, but not MKK1a and MKK2a, could bypass the requirement of FLS2 in flg22 signalling, demonstrated by their ability to activate the *WRKY29* and *FRK1* promoters in *fls2* mutant protoplasts (Fig. 3e).

Thus, on the basis of both gain-of-function and loss-of-function analyses, it appears that flagellin signalling activates MKK4 and MKK5, which in turn phosphorylate and activate MPK3 and MPK6, leading to the expression of early-defence response genes. The functions of MKK4 and MKK5 in flg22 signalling are probably redundant in *Arabidopsis* leaf cells (see below). It has been proposed that MPK3 and/or MPK6 mediate expression of *GST6* in *Arabidopsis* protoplasts treated with H<sub>2</sub>O<sub>2</sub> (ref. 35). A MAPKK functioning in this oxidative-stress signalling pathway has not been identified. Although MKK4 and MKK5 can activate MPK3 and MPK6, neither MKK4a nor MKK5a was able to induce the expression of *GST6*, a reporter gene in the oxidative-stress signalling pathway (Fig. 3c, e). This result suggests that flagellin and H<sub>2</sub>O<sub>2</sub> may activate different MAPK signalling cascades.

### A specific MAPKKK in flg22 signalling

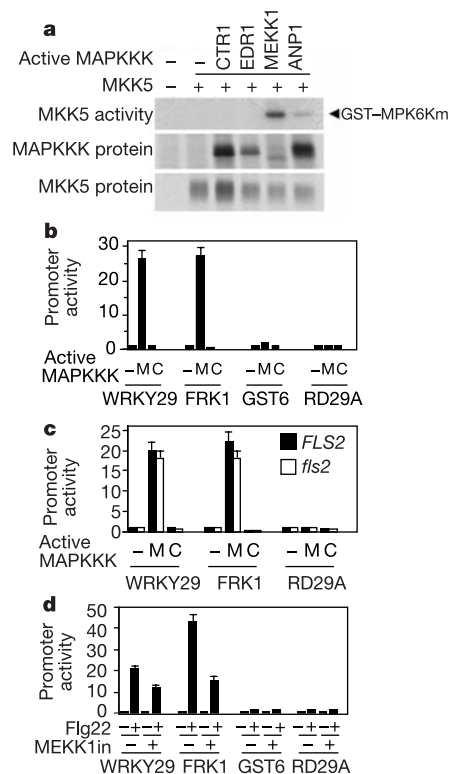
On the basis of sequence homology in the kinase domains to mammalian MAPKKKs, there are at least 25 putative *Arabidopsis* MAPKKKs that can be divided into six subgroups<sup>19</sup>. The sequence similarity among the subgroups is low, especially in the regulatory domains, indicating that the subgroups have distinct functions<sup>19,38</sup>. We analysed four *Arabidopsis* MAPKKKs (CTR1, EDR1, MEKK1 and ANP1; refs 19, 38) from three subgroups expressed in leaf cells, to determine whether any of them could activate MKK5. Constitutively active derivatives of these four MAPKKKs were constructed by deleting the putative regulatory domain<sup>35</sup>. The four constitutively active MAPKKKs were fused to a HA-epitope tag and co-expressed individually with Myc-epitope-tagged wild-type MKK5 in protoplasts. To determine whether MKK5 had been phosphorylated and activated *in vivo*, MKK5 was immunoprecipitated using an anti-Myc antibody, and its activity was determined *in vitro* using a purified kinase-inactive MPK6 fused to glutathione S-transferase (GST-MPK6Km) as a substrate. As shown in Fig. 4a, MKK5 was specifically activated by constitutively active MEKK1 ( $\Delta$ MEKK1) and could phosphorylate MPK6. Constitutively active ANP1 ( $\Delta$ ANP1), which can mimic the H<sub>2</sub>O<sub>2</sub> signal and induce the *GST6* promoter<sup>35</sup>, activated MKK5 only marginally despite a high level of protein expression (Fig. 4a), suggesting that flagellin and H<sub>2</sub>O<sub>2</sub> activate different MAPK pathways. A control experiment using a kinase-inactive mutant of MKK5 ruled out the possibility of a direct phosphorylation of GST-MPK6Km by MEKK1 (not shown).

Consistent with the above results,  $\Delta$ MEKK1 activated the *WRKY29* and *FRK1* promoters in the absence of flg22 in both wild-type and *fls2* mutant protoplasts (Fig. 4b, c) but not the activity of the H<sub>2</sub>O<sub>2</sub>- or  $\Delta$ ANP1-inducible *GST6* promoter (Fig. 4b). Constitutively active CTR1 ( $\Delta$ CTR1), which did not activate MKK5, had no effect on the *WRKY29* and *FRK1* promoter activities (Fig. 4b, c). To test further the involvement of MEKK1 in flg22 signalling, a dominant-negative mutant was generated by altering the ATP-binding site but maintaining the regulatory and protein-protein interaction domains of the full-length MEKK1. The dominant-negative mutant (MEKK1in) partially suppressed flg22 activation of *WRKY29* and *FRK1* promoters (Fig. 4d), similarly to U0126 treatment, MKP1, and dominant-negative MKK4 or MKK5 (Figs 1b, d and 3d). Dominant-negative CTR1 did not have

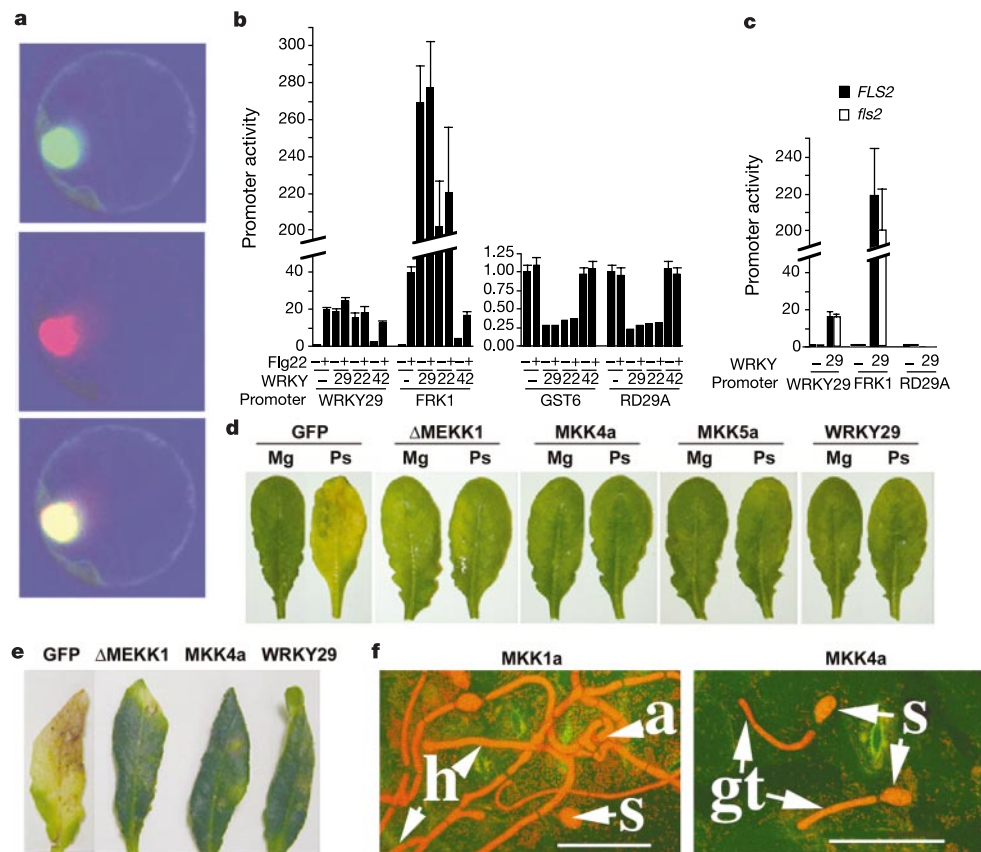
an effect (not shown). Taken together, the results described so far strongly suggest that flagellin perception by FLS2 leads to the induction of *WRKY29* and *FRK1* transcription through the activation of a MAPK signalling cascade consisting of MEKK1, MKK4/MKK5, and MPK3/MPK6.

### Positive feedback control by WRKYs

WRKY proteins have been shown to bind to W-box DNA elements (containing a TGAC core sequence) that are found in the promoters of many defence-related genes including *WRKY29* (10 W boxes) and *FRK1* (15 W boxes)<sup>28,32–34</sup>. In parsley protoplasts, a fungal pathogen-inducible WRKY1 protein, with two WRKY domains, is targeted to the nucleus and activates its own promoter by binding to multiple W boxes<sup>33</sup>. *Arabidopsis* WRKY29, with a single WRKY domain, is not the orthologue of parsley WRKY1 (ref. 33). WRKY29-GFP, a WRKY29 and green fluorescent protein fusion, was constitutively localized in the nucleus in the presence or absence of flg22 (Fig. 5a). Transient expression of WRKY29 activated its own promoter in the absence of flg22 (Fig. 5b), indicating a positive feedback control<sup>33</sup>. WRKY29 also strongly activated the *FRK1* promoter, but suppressed the basal activities of the *GST6* and *RD29A* promoters (Fig. 5b). In the *fls2* mutant protoplasts, transient expression of WRKY29 activated the *WRKY29* and *FRK1*, but not *RD29A*, promoters, suggesting that WRKY29 acts downstream of the flagellin receptor (Fig. 5c). Treatment of WRKY29-expressing protoplasts with flg22 did not further enhance the activity of the *WRKY29* and *FRK1* promoters (Fig. 5b). These results are consistent



**Figure 4** MEKK1 initiates the flg22 MAPK cascade. **a**, Constitutively active MEKK1 activates MKK5. Protoplasts were expressing MKK5 and one constitutively active MAPKKK. MKK5 activity using GST-MPK6Km as a substrate (top) and MAPKKK (middle) and MKK5 (bottom) expression are shown. **b**, Constitutively active MEKK1 induces the *WRKY29* and *FRK1* promoters. Protoplasts were expressing GFP or constitutively active MEKK1 (M) or CTR1 (C). **c**, MEKK1 acts downstream of FLS2. Protoplasts were isolated from wild-type (*FLS2*) and mutant (*fls2*) leaves. **d**, Dominant-negative MEKK1 inhibits flg22 activation of the *WRKY29* and *FRK1* promoters. Protoplasts were expressing a kinase-inactive MEKK1 (MEKK1in).



**Figure 5** The flg22 MAPK cascade and specific WRKYs are important for *Arabidopsis* defence. **a**, Nuclear localization of WRKY29-GFP. Visualization of protoplasts expressing WRKY29-GFP (top), a red-fluorescent nuclear marker (middle), and super-imposition (bottom) without flg22. **b**, WRKY22 and WRKY29 activate early-defence genes. Protoplasts were expressing GFP, WRKY29 (29), WRKY22 (22) or WRKY42 (42). **c**, WRKY29 acts downstream of FLS2. Protoplasts were isolated from wild-type (*FLS2*) and mutant (*fls2*) leaves. **d**, Leaves expressing  $\Delta$ MEKK1, MKK4a, MKK5a or WRKY29

exhibit reduced disease symptoms after *P. syringae* (Ps) infection. Control was infiltrated with 10 mM  $\text{MgSO}_4$  (Mg). **e**, Leaves expressing  $\Delta$ MEKK1, MKK4a or WRKY29 exhibit reduced disease symptoms after *B. cinerea* infection on the right half of each leaf. **f**, The early stage of *B. cinerea* development was inhibited on leaves expressing MKK4a. On MKK1a-expressing leaves (MKK1a), germinated spores (s) of *B. cinerea* formed superficial hyphae (h) and branched appressoria (a) 2 days after infection. The fungal spores (s) formed only germ tubes (gt) on MKK4a-expressing leaves (MKK4a). Scale bars, 50  $\mu\text{m}$ .

with the idea that unmodified wild-type WRKY29 can act as a transcriptional activator. In contrast to WRKY29, a different *Arabidopsis* WRKY protein, WRKY42 (At4g04450), did not activate the *WRKY29* and *FRK1* promoters, but instead slightly inhibited the flg22 activation of these promoters (Fig. 5b). Moreover, unlike WRKY29, WRKY42 did not suppress basal activities of the *GST6* and *RD29A* promoters (Fig. 5b). Thus, we suggest that WRKY29 is a key transcriptional activator involved in the expression of defence genes in *Arabidopsis* innate immune responses. Similarly to MKK and MPK gene families, the large WRKY family also has members that are highly homologous to WRKY29 (ref. 32). The homologous WRKY22 (At4g01250) in the same WRKY subgroup<sup>32</sup> was cloned and tested in the protoplast transient assay. WRKY22 regulated these promoters similarly to WRKY29 (Fig. 5b), suggesting that WRKY29 and WRKY22 may be functionally redundant.

### Resistance to bacterial and fungal pathogens

On the basis of the results obtained with the protoplast transient expression assays, we tested whether *Arabidopsis* plants exhibited enhanced resistance to bacterial pathogens when the flagellin MAPK cascade is constitutively activated or WRKY29 is expressed by means of *Agrobacterium*-mediated transient transformation<sup>39</sup>. When infected by the virulent bacterial pathogen *Pseudomonas syringae*, transiently transformed control leaves expressing GFP developed chlorotic lesions (Fig. 5d). Similar results were obtained with untransformed leaves and leaves expressing  $\Delta$ CTR1, MKK1a

or wild-type WRKY42 as controls (not shown). In contrast, transformed leaves expressing  $\Delta$ MEKK1, MKK4a, MKK5a or wild-type WRKY29 displayed enhanced resistance to *P. syringae* (Fig. 5d). The observation that both MKK4a and MKK5a can confer pathogen resistance is consistent with their functional redundancy. We note that development of soft-rot symptoms caused by infection with the fungal pathogen *Botrytis cinerea* was also effectively suppressed when  $\Delta$ MEKK1, MKK4a or WRKY29, but not GFP, was transiently expressed in leaves (Fig. 5e). Microscopic analysis indicated that in *Arabidopsis* leaves expressing MKK1a as a control, fungal spores germinated and formed superficial mycelium and branched appressoria filled with intensely stained cytoplasm within two days after *B. cinerea* infection (Fig. 5f). In contrast, only germ tubes were formed during the same period of time in leaves expressing MKK4a. Similar results were obtained with  $\Delta$ MEKK1, MKK5a and WRKY29 (not shown). These results suggest that defence responses activated by the flagellin MAPK cascade or WRKY29 are effective against both fungal and bacterial pathogens.

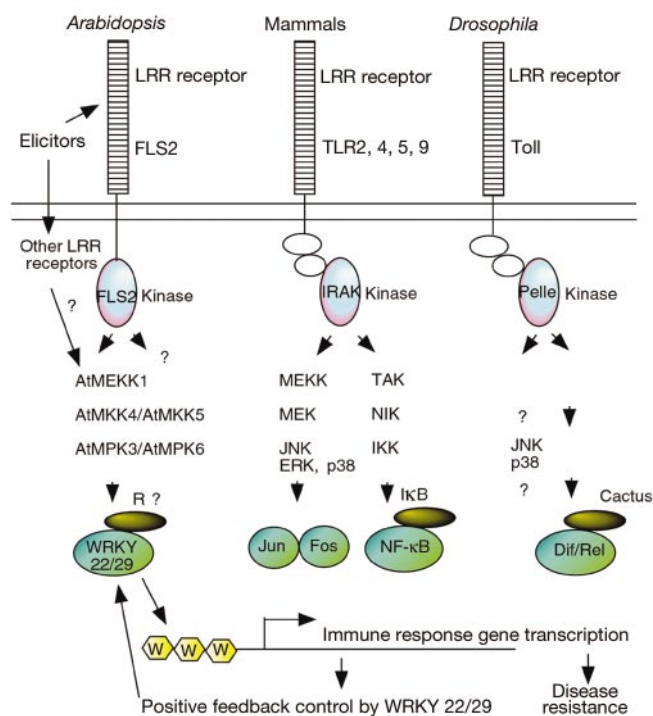
### Discussion

We have demonstrated that *Arabidopsis* mesophyll protoplast transient expression assays can be combined with genetic and genomic information to provide a powerful tool for the analysis of MAPK signalling involved in plant innate immunity. Because wild-type WRKY22 or WRKY29 is sufficient to mimic flg22 and MAPK signalling, we suggest that a specific WRKY inhibitor may be



phosphorylated and inactivated by the flagellin MAPK cascade upon pathogen infection. As summarized in the model shown in Fig. 6, this would make the flg22 signalling pathway partially analogous to signalling pathways in insects and mammals, in which Toll and TLRs, respectively, initiate a cascade of signalling events leading to phosphorylation and degradation of an inhibitor of a Rel-like transcription factor<sup>2,4</sup>. It will be interesting to test whether the flg22- and FLS2-dependent phosphorylation of the ankyrin-containing protein AtPhos43, recently identified by a proteomic approach<sup>40</sup>, is MAPK dependent and can control WRKY22 and WRKY29 subcellular localization and/or activity.

The flagellin MAPK cascade leading to the activation of *Arabidopsis* MPK3 and MPK6 is reminiscent of the activation of the tobacco orthologues of MPK3 and MPK6, WIPK and SIPK, respectively, by fungal elicitor, as well as during Avr9-Cf-9 and TMV-N gene-for-gene interactions<sup>14,21,22</sup>. Recently, Yang *et al.*<sup>41</sup> identified a tobacco MAPKK, NtMEK2, which can activate WIPK and SIPK. Expression of a constitutively active mutant of NtMEK2 in tobacco leaves led to the induction of hypersensitive cell death and the expression of defence genes in the absence of pathogens. The results suggest that a MAPK cascade containing NtMEK2, WIPK and SIPK is involved in the expression of fungal pathogen defence responses in tobacco<sup>41</sup>. The *Arabidopsis* orthologues of NtMEK2 are MKK4 and MKK5<sup>19,41</sup>, further suggesting the importance of the flagellin MAPK cascade in pathogen defence. Moreover, the data in Fig. 5 suggest that the *Arabidopsis* flagellin MAPK cascade may also be involved in expression of fungal defence responses without eliciting extensive hypersensitive cell death as in the case of NtMEK2 in tobacco. This, in turn, suggests that signalling events initiated by diverse pathogens converge into a conserved MAPK cascade (shown in Fig. 6)<sup>7,14,17–19,23,41</sup>. However, additional quantitative analyses in both transiently and stably transformed plants are required to confirm the proposed role of the flg22 MAPK defence-response pathway.



**Figure 6** Model of innate immune signalling activated by LRR receptors in *Arabidopsis*, mammals and *Drosophila*. A putative repressor (R) could control WRKY22 and WRKY29 activity because their overexpression bypasses the requirement of elicitors. The conserved signalling pathways for innate immune responses in animals are summarized on the basis of recent reviews on mammals<sup>2</sup> and *Drosophila*<sup>4</sup>.

As shown in Figs 1b, d, 3d and 4d, our data also reveal the previously unexpected existence of a MAPK-independent pathway in flg22 signalling. Characterization of this new pathway, which appears to be as important as the MAPK pathway for the induction of defence-related genes, will probably uncover additional complexity in the signalling pathways that control the expression of early-defence-related genes. Future work may also allow these MAPK-dependent and MAPK-independent pathways to be linked with other defence signalling components, such as NDR1, EDS1, PAD4, EDR1, MPK4, PBS1 and PBS2, defined by *Arabidopsis* mutants<sup>8,31,42–44</sup>.

The lack of MAPK cascade mutations in the flg22 signalling pathway is most probably due to functional redundancy in the pathway, as suggested by the identification of the MKK4/MKK5 and MPK3/MPK6 pairs in our study. The WRKY22 and WRKY29 transcription factors in the flg22 signalling pathway may also provide redundant functions. Furthermore, similar to the yeast MAPKKK STE11 (ref. 45), some plant MAPK signalling components could be involved in more than one pathway<sup>19</sup>. The H<sub>2</sub>O<sub>2</sub>-activated MAPK cascade shares MPK3 and MPK6 with the flg22 pathway but activates different target genes<sup>35</sup>. The MAPKKK of the flagellin cascade, MEKK1, may also initiate a different MAPK cascade. On the basis of two-hybrid analyses and mutant complementation assays in yeast, an *Arabidopsis* MAPK cascade (MEKK1-MKK1/MKK2-MPK4) has been proposed<sup>37</sup>. In addition, we note that, besides the flagellin MAPK cascade, other pathways using untested MAPK, MAPKK and MAPKKK genes could also be important in *Arabidopsis* innate immunity. Finally, plants transiently expressing ΔMEKK1, MKK4a or WRKY29 exhibited enhanced resistance to both bacterial and fungal pathogens (Fig. 5), suggesting that other pathogen-derived signals in addition to flg22 may activate the MEKK1, MKK4/MKK5, MPK3/MPK6, WRKY22/WRKY29 signalling pathway through distinct receptors (Fig. 6)<sup>40</sup>. Given this complexity, a combination of genetic, genomic, cellular and biochemical approaches including the construction of knockout mutants and expression of signalling components in transgenic plants will be required to untangle the intertwined signalling webs mediated by MAPK cascades in plants.

**Note added in proof:** Transient expression of MKK4a in *Arabidopsis* leaves, as in Fig. 5d, inhibited the growth of *P. syringae* maculicola ES4326 15-fold one day after infiltration, and over 100-fold four days after infiltration. The results are consistent with the leaf phenotype shown in Fig. 5d. □

## Methods

### Reporter constructs

To identify early-response genes, a subtracted cDNA library that represented mRNA species induced by the elicitor flg22 in *Arabidopsis* mesophyll protoplasts was generated as previously described<sup>30</sup>. The DNA regions immediately upstream from the translation start sites of the WRKY29, FRK1 and GST1 genes were amplified by PCR from *Arabidopsis* (Col-0) genomic DNA. The sizes of the amplified fragments were 2.6 kilobases (kb), 2.8 kb, and 0.9 kb, respectively. The promoters were fused to a luciferase reporter gene to generate WRKY29-LUC, FRK1-LUC and GST1-LUC<sup>26,35</sup>.

### Effector constructs

The *Arabidopsis* cDNAs, encoding MAPKs, were: MPK2 (Atlg59580), MPK3 (At3g45640), MPK5 (At4g11330), MPK6 (At2g43790), MPK7 (At2g18170), MPK9 (At3g18040). The MAPKKs were: MKK1 (At4g26070), MKK2 (At4g29810), MKK4 (Atlg51660), MKK5 (At3g21220). The MAPKKKs were: CTR1 (At5g03730), EDR1 (Atlg08720), MEKK1 (At4g08500), ANP1 (Atlg09000). Others were receptor FLS2 (At5g46330), transcription factors WRKY22 (At4g01250), WRKY29 (At4g23550) and WRKY42 (At4g04450). All cDNAs were obtained by PCR from an *Arabidopsis* cDNA library<sup>26,35</sup> and verified by sequencing. Mouse MKP1 was a gift from H. Sun and N. Tonks<sup>36</sup>. PCR products were fused to the double HA (MAPK, MAPKK, receptor and WRKY) or Myc (MAPKK and MKP1) epitope tag sequence and inserted into a plant expression vector or pCB302 containing the 35S4PPDK promoter and the NOS terminator<sup>35,36,39</sup>.

The constitutively active forms of MAPKKs were generated by site-specific mutation, replacing the serine or threonine residues in the activation loop domain [S/T]XXXX[S/T] by acidic amino acids glutamate or aspartate: MKK1(T218ES224D), MKK2(T220DT226E), MKK4(T224DS230E) and MKK5(T215ES221E). The active forms of MAPKKK were generated by keeping the catalytic domain only: ΔCTR1(544–799),

$\Delta$ EDR1(662–921),  $\Delta$ MEKK1(326–592),  $\Delta$ ANP1(57–338). The inactive forms of all kinases were generated by site-specific mutation, replacing the conserved lysine residues in the kinase ATP-binding loop by a methionine: FLS2(K898M), MKK4(K108M), MKK5(K99M), MPK6(K92M,K93M), MEKK1(K361M). All these mutations were verified by sequencing.

# Arabidopsis mesophyll protoplast transient expression assay

Protoplast transient expression assay was carried out as described previously<sup>26,35</sup>. Promoter activities are represented by LUC/GUS activities, and normalized to the value obtained with protoplasts without the treatments. In case of protoplasts prepared from *fls2* mutant plants, promoter activities are normalized to the value obtained with wild-type protoplasts transfected with a control GFP plasmid. Constitutively active MAPKKKs, immunoprecipitated from transfected protoplasts, phosphorylated casein *in vitro*<sup>36</sup>, indicating that they are active kinases. GFP fluorescence was observed by either Nikon TE200 fluorescent microscopy or with Leica TCS NT Confocal Spectrophotometer (Germany). Co-expression of a nuclear-targeted red-fluorescent protein was used as a control<sup>46</sup>.

# RT-PCR

Samples were taken at indicated time points after treatment and total RNA was prepared as described<sup>30</sup>. RT-PCR was performed with 1 ng of total RNA and 0.6  $\mu$ M of each primer using OneStep RT-PCR Kit (Qiagen). PCR was run for 35 cycles. Constitutively expressed *UBQ10* (AT4g05320) or *ACTIN-1* (At2g37620) mRNA was always co-amplified with each mRNA and used as an internal standard.

# Protein kinase assays

The MAPK in-gel kinase assay was carried out as described<sup>35</sup>. For immuno-complex assays, tagged kinases were immunoprecipitated from lysates of transfected protoplasts with the corresponding antibody and analysed with a known substrate as indicated. The GST–MPK6Km fusion protein was generated by subcloning the coding sequence into pGEX-4T-1 vector (Amersham Pharmacia Biotech) in frame with the GST coding sequence. The fusion protein was purified according to the manufacturer's procedure and used as a substrate for *in vitro* phosphorylation assay. Except for the kinase-inactive mutants that were used as controls, all protein kinases showed significant auto-phosphorylation activity and were able to phosphorylate an exogenous substrate, indicating that they are active kinases (not shown).

# Agrobacterium-mediated transient transformation

To analyse plant susceptibility to a bacterial pathogen, five-week-old *Arabidopsis* plants were infiltrated with a suspension (containing 100  $\mu$ M acetosyringone) of *A. tumefaciens* carrying a binary vector pCB302 (ref. 39) expressing GFP,  $\Delta$ MEKK1, MKK4a, MKK5a or WRKY29. The plants were incubated for three days under the conditions described<sup>30</sup>. Infiltrated leaves were then inoculated with 10 mM MgSO<sub>4</sub> or the same solution with suspended *P. syringae* *maculicola* ES4326 (10<sup>4</sup> colony-forming units, CFU, per cm<sup>2</sup>) at the location of transient gene expression. The leaves were photographed three days later.

To analyse plant susceptibility to a fungal pathogen, a suspension (containing 100  $\mu$ M acetosyringone and 0.01% Silwett L-77) of *A. tumefaciens* expressing GFP,  $\Delta$ MEKK1, MKK4a, MKK1a or WRKY29 was first applied to the lower surface of a leaf of six-week-old *Arabidopsis*. The leaves were incubated for 24 h and then the upper surface of transformed leaves was infected with *B. cinerea* sclerotia. The infected leaves were photographed five days later after removing the fungal sclerotia. To observe the early stage of *B. cinerea* development, 5- $\mu$ l drops of a  $5 \times 10^5$  spores ml<sup>-1</sup> suspension were placed on the upper surface of an *Arabidopsis* leaf. Two days after infection, fungal structures were visualized by trypan blue lactophenol staining and observed with Confocal Spectrophotometer TCS NT (Leica, Germany). Untransformed and GFP-expressing control leaves developed similar lesions to those on the MKK1a-expressing leaves.

Received 28 August; accepted 24 December 2001.

- Boller, T. Chemoperception of microbial signals in plant cells. *Annu. Rev. Plant Physiol. Plant. Mol. Biol.* **46**, 189–214 (1995).
- Aderem, A. & Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782–787 (2000).
- Meindl, T., Boller, T. & Felix, G. The bacterial elicitor flagellin activates its receptor in tomato cells according to the address-message concept. *Plant Cell* **12**, 1783–1794 (2000).
- Khush, R. S. & Lemaitre, B. Genes that fight infection. *Trends Genet.* **16**, 442–449 (2000).
- Cohn, J., Sessa, G. & Martin, G. B. Innate immunity in plants. *Curr. Opin. Immunol.* **13**, 55–62 (2001).
- Dangl, J. L. & Jones, J. D. G. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833 (2001).
- Nurnberger, T. & Scheel, D. Signal transmission in plant immune response. *Trends Plant Sci.* **6**, 372–379 (2001).
- Staskawicz, B. J., Mudgett, M. B., Dangl, J. L. & Galan, J. E. Common and contrasting themes of plant and animal diseases. *Science* **292**, 2285–2289 (2001).
- Samakovlis, C., Asling, B., Boman, H. G., Gateff, E. & Hultmark, D. *In vitro* induction of cecropin genes—an immune response in a *Drosophila* blood cell line. *Biochem. Biophys. Res. Commun.* **188**, 1169–1175 (1992).
- Felix, G., Duran, J., Volko, S. & Boller, T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276 (1999).
- Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
- Gomez-Gomez, L. & Boller, T. FLS2: An LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol. Cell* **5**, 1003–1011 (2000).
- Gomez-Gomez, L., Felix, G. & Boller, T. A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *Plant J.* **18**, 277–284 (1999).

- Romeis, T. *et al.* Rapid *Avr9*- and *Cf9*-dependent activation of MAP kinases in tobacco cell cultures and leaves: Convergence of resistance gene, elicitor, wound, and salicylate responses. *Plant Cell* **11**, 273–287 (1999).
- Blume, B., Nurnberger, T., Nass, N. & Scheel, D. Receptor-mediated increase in cytoplasmic free calcium required for activation of pathogen defense in parsley. *Plant Cell* **12**, 1425–1440 (2000).
- Hirt, H. & Scheel, D. in *Results and Problems in Cell Differentiation: MAP Kinases in Plant Signal Transduction* (ed. Hirt, H.) 85–93 (Springer, Heidelberg, 2000).
- Zhang, S. & Klessig, D. F. MAPK cascades in plant defense signaling. *Trends Plant Sci.* **6**, 520–527 (2001).
- Romeis, T. Protein kinases in the plant defence response. *Curr. Opin. Plant Biol.* **4**, 407–414 (2001).
- Tena, G., Asai, T., Chiu, W.-L. & Sheen, J. Plant MAP kinase signaling cascades. *Curr. Opin. Plant Biol.* **4**, 392–400 (2001).
- Ligterink, W., Kroj, T., Zurnieden, U., Hirt, H. & Scheel, D. Receptor-mediated activation of a MAP kinase in pathogen defense in plants. *Science* **276**, 2054–2057 (1997).
- Zhang, S. & Klessig, D. F. Resistance gene *N*-mediated *de novo* synthesis and activation of a tobacco mitogen-activated protein kinase by tobacco mosaic virus infection. *Proc. Natl Acad. Sci. USA* **95**, 7433–7438 (1998).
- Zhang, S., Du, H. & Klessig, D. F. Activation of the tobacco SIP kinase by both a cell-wall-derived carbohydrate elicitor and purified proteinaceous elicitors from *Phytophthora* spp. *Plant Cell* **10**, 435–449 (1998).
- Nuhse, T. S., Peck, S. C., Hirt, H. & Boller, T. Microbial elicitors induce activation and dual phosphorylation of the *Arabidopsis thaliana* MAPK6. *J. Biol. Chem.* **275**, 7521–7526 (2000).
- Cardinale, F. *et al.* Differential activation of four specific MAPK pathways by distinct elicitors. *J. Biol. Chem.* **275**, 36734–36740 (2000).
- Lee, J., Klessig, D. F. & Nurnberger, T. A. A harpin binding site in tobacco plasma membranes mediates activation of the pathogenesis-related gene *HIN1* independent of extracellular calcium but dependent on mitogen-activated protein kinase activity. *Plant Cell* **13**, 1079–1093 (2001).
- Sheen, J. Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. *Plant Physiol.* (in the press).
- Durrant, W. E., Rowland, O., Piedras, P., Hammond-Kosack, K. E. & Jones, J. D. G. cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. *Plant Cell* **12**, 963–977 (2000).
- Maleck, K. *et al.* The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genet.* **26**, 403–410 (2000).
- Schenk, P. M. *et al.* Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl Acad. Sci. USA* **97**, 11655–11660 (2000).
- Asai, T. *et al.* Fumonisin B1-induced cell death in *Arabidopsis* protoplasts requires jasmonate-, ethylene-, and salicylate-dependent signaling pathways. *Plant Cell* **12**, 1823–1835 (2000).
- Glazebrook, J. Genes controlling expression of defense responses in *Arabidopsis*—2001 status. *Curr. Opin. Plant Biol.* **4**, 301–308 (2001).
- Eulgem, T., Rushton, P. J., Robatzek, S. & Somssich, I. E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**, 199–206 (2000).
- Eulgem, T., Rushton, P. J., Schmeizer, E., Hahlbrock, K. & Somssich, I. E. Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors. *EMBO J.* **18**, 4689–4699 (1999).
- Du, L. & Chen, Z. Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis*. *Plant J.* **24**, 837–847 (2000).
- Kovtun, Y., Chiu, W.-L., Tena, G. & Sheen, J. Function analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl Acad. Sci. USA* **97**, 2940–2945 (2000).
- Kovtun, Y., Chiu, W.-L., Zeng, W. & Sheen, J. Suppression of auxin signal transduction by a MAPK cascade in higher plants. *Nature* **395**, 716–720 (1998).
- Mizoguchi, T., Ichimura, K., Yoshida, R. & Shinozaki, K. in *Results and Problems in Cell Differentiation: MAP Kinases in Plant Signal Transduction* (ed. Hirt, H.) 29–38 (Springer, Heidelberg, 2000).
- Jouanin, S. *et al.* Plant MAP kinase kinase kinase structure, classification and evolution. *Gene* **233**, 1–11 (1999).
- Xiang, C., Han, P., Lutziger, L., Wang, K. & Oliver, D. J. A mini binary vector series for plant transformation. *Plant Mol. Biol.* **40**, 711–717 (1999).
- Peck, S. C. *et al.* Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* **13**, 1467–1475 (2001).
- Yang, K.-Y., Liu, Y. & Zhang, S. Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco. *Proc. Natl Acad. Sci. USA* **98**, 741–746 (2001).
- Petersen, M. *et al.* *Arabidopsis* MAP kinase 4 negatively regulates systemic acquired resistance. *Cell* **103**, 1111–1120 (2000).
- Frye, C. A., Tang, D. & Innes, R. W. Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proc. Natl Acad. Sci. USA* **98**, 373–378 (2001).
- Swiderski, M. R. & Innes, R. W. The *Arabidopsis* *PBS1* resistance gene encodes a member of a novel protein kinase subfamily. *Plant J.* **26**, 101–112 (2001).
- Madhani, H. D. & Fink, G. R. The riddle of MAP kinase signalling specificity. *Trends Genet.* **14**, 151–155 (1998).
- Patharkar, O. R. & Cushman, J. C. A stress-induced calcium-dependent protein kinase from *Mesembryanthemum crystallinum* phosphorylates a two-component pseudo-response regulator. *Plant J.* **24**, 679–691 (2000).

# Acknowledgements

We thank C. Xiang for the pCB302 minibinary vector, O. R. Patharkar and J. Cushman for the DsRed-CSP1 plasmid; S. Volko and J. Stone for sharing unpublished information; S.-H. Cheng, B. Moore and I. Hwang for technical advice; W.-C. Chen for help in bioinformatics; and S. Ramu and F. Rolland for critical reading of the manuscript. This work was supported by a NSF plant genome project grant to J.S. and F.M.A., a USDA grant to J.S., and a NIH grant to F.M.A. T.A. was supported in part by fellowships from the Toyobo Biotechnology Foundation and the Uehara Memorial Foundation.

Correspondence and requests for materials should be addressed to J.S. (e-mail: sheen@molbio.mgh.harvard.edu).

# Control of Jupiter's radio emission and aurorae by the solar wind

D. A. Gurnett\*, W. S. Kurth\*, G. B. Hospodarsky\*, A. M. Persoon\*, P. Zarka†, A. Lecacheux‡, S. J. Bolton‡, M. D. Desch§, W. M. Farrell§, M. L. Kaiser§, H.-P. Ladreiter||, H. O. Rucker||, P. Galopeau¶, P. Louarn#, D. T. Young\*, W. R. Pryor\*\* & M. K. Dougherty††

\* Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

† Space Research Department, Observatoire de Paris, 92195 Meudon, France

‡ Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

§ NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

|| Space Research Institute, Austrian Academy of Science, A-8010 Graz, Austria

¶ Centre d'étude des Environnements, Terrestre et Planétaires, Université de Versailles Saint-Quentin-en-Yvelines (CETP/UVSQ), 10–12 Avenue de l'Europe, F-78140 Velizy, France

# Centre d'Etude Spatiale des Rayonnements (CESR-CNRS), 9 Avenue du Colonel Roche, Centre National de la Recherche Scientifique, 31028 Toulouse, France

\* Department of Atmospheric, Oceanic and Space Science, University of Michigan, Ann Arbor, Michigan 48109, USA

\*\* Laboratory for Atmospheric and Space Physics (LASP), University of Colorado, Boulder, Colorado 80303, USA

†† Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

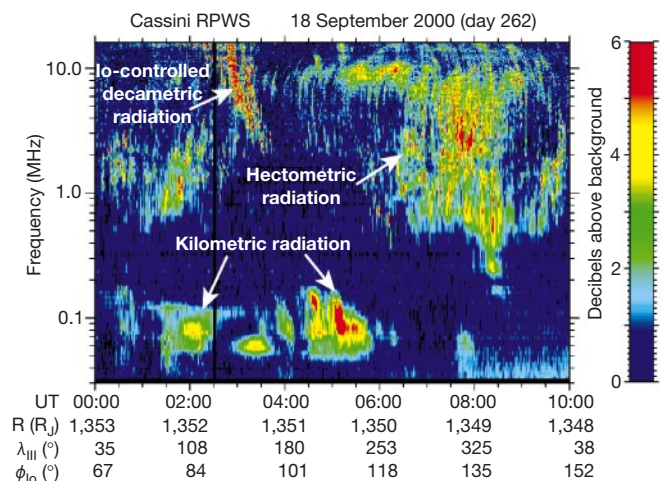
Radio emissions from Jupiter provided the first evidence that this giant planet has a strong magnetic field<sup>1,2</sup> and a large magnetosphere<sup>3</sup>. Jupiter also has polar aurorae<sup>4</sup>, which are similar in many respects to Earth's aurorae<sup>5</sup>. The radio emissions are believed to be generated along the high-latitude magnetic field lines by the same electrons that produce the aurorae, and both the radio emission in the hectometric frequency range and the aurorae vary considerably<sup>6,7</sup>. The origin of the variability, however, has been poorly understood. Here we report simultaneous observations using the Cassini and Galileo spacecraft of hectometric radio emissions and extreme ultraviolet auroral emissions from Jupiter. Our results show that both of these emissions are triggered by interplanetary shocks propagating outward from the Sun. When such a shock arrives at Jupiter, it seems to cause a major compression and reconfiguration of the magnetosphere, which produces strong electric fields and therefore electron acceleration along the auroral field lines, similar to the processes that occur during geomagnetic storms at the Earth.

The Cassini fly-by of Jupiter on 30 December 2000 provided an opportunity to carry out coordinated measurements between the Cassini spacecraft, which is on its way to Saturn, and the Galileo spacecraft, which has been orbiting Jupiter since 1995. Figure 1 shows a frequency–time spectrogram obtained from the radio and plasma wave science (RPWS) instrument<sup>8</sup> on Cassini during the approach to Jupiter. Several different types of radio emissions can be seen in this spectrogram, but the most prominent are the distinct arc-like features from about 0.5 to 16 MHz. These arc-like features were first clearly seen during the Voyager 1 fly-by of Jupiter<sup>9</sup>, and are believed to be generated at the local electron cyclotron frequency via the cyclotron maser mechanism<sup>10</sup>, which produces radiation nearly perpendicular to the local magnetic field. Usually, two types of arcs can be identified, those associated with Jupiter's moon Io<sup>11</sup>, labelled 'Io-controlled', and those not associated with Io. The Io-controlled arcs are strongest in the decametric frequency range, and are produced by a system of Alfvén waves excited by Io<sup>12,13</sup>. The non-Io-controlled emissions are independent of Io's position and are highly variable. These emissions are often called hectometric radiation, because they mainly occur at hectometric wavelengths

(0.3–3 MHz), although some radiation also extends upward into the decametric frequency range (3–30 MHz). Previous studies have shown that the hectometric radiation is generated at altitudes of several jovian radii along the high-latitude auroral field lines<sup>14</sup>.

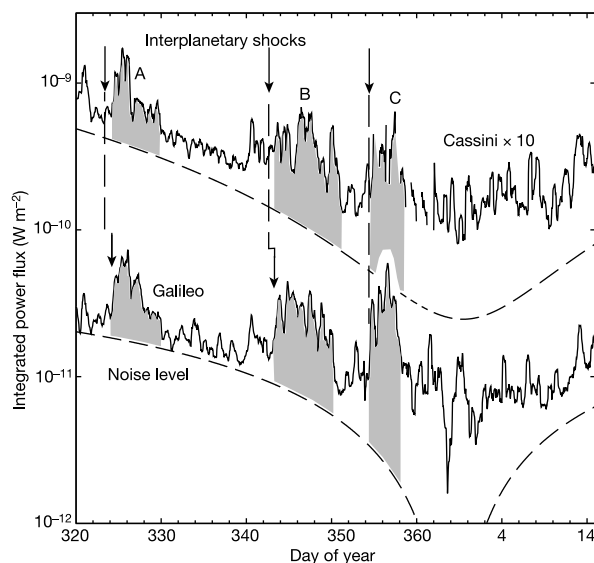
In this study we have concentrated on investigating the solar wind control of jovian hectometric radiation and the auroral extreme ultraviolet emissions. Evidence already exists that the solar wind plays a role in controlling the intensity of the hectometric radiation<sup>15,16</sup>. To provide a parameter that characterizes the intensity of the jovian hectometric radiation, we have integrated the radio emission spectrums detected by the Cassini RPWS instrument and the Galileo plasma wave science (PWS)<sup>17</sup> instrument over a frequency range from 0.5 to 5.6 MHz. The corresponding radiated power fluxes, in  $\text{W m}^{-2}$ , are shown in Fig. 2: large variations are present in the hectometric radio emission intensity. Three well defined periods of enhanced hectometric radiation are evident. These events are indicated by the shading, and are labelled events A, B and C. The magnetometer and plasma measurements on Cassini show that each of these events is preceded by an interplanetary shock. The arrival times of the shocks at Cassini are indicated by the arrows at the top of the plot.

Further details of event A are shown on an expanded timescale in Fig. 3. The arrival of the interplanetary shock at Cassini is clearly indicated by the jump in the magnetic field at the arrow marked 'shock' in Fig. 3a. The shock is followed by a region of strongly enhanced magnetic field lasting about 5 days, with a peak on day 326. Figure 3b shows that the ion density increased markedly over the same period, indicating that the shock is followed by a region of highly compressed plasma. Such regions of compressed plasma are typical of interplanetary shocks at this distance from the Sun. The same interplanetary shock was subsequently detected by the Galileo plasma wave instrument near Jupiter about 16 h later at 09:15 UT on day 324. From the approximately 16-h travel time from Cassini (at  $x = +540R_J$ , where  $x$  is the distance toward the Sun from Jupiter and  $R_J$  is the radius of Jupiter) to Galileo (at  $x = -52R_J$ ), the average shock propagation speed is estimated to be  $721 \text{ km s}^{-1}$ . This speed is typical of a relatively strong interplanetary shock.



**Figure 1** A representative frequency–time spectrogram of jovian radio emissions detected by the Cassini spacecraft during the approach to Jupiter at a radial distance of about 1,350 Jupiter radii ( $R_J$ ). The colour bar on the right gives the radio emission intensities in decibels above the background noise level, which is approximately the cosmic background. UT, universal time at the spacecraft, in hours and minutes.  $R$ , the radial distance of the spacecraft from Jupiter in  $R_J$ .  $\lambda_{III}$ , the system III longitude of the spacecraft.  $\phi_{Io}$ , the phase of Io relative to the spacecraft. Three types of jovian radio emissions are evident in this spectrogram: Io-controlled decametric radiation, hectometric radiation and kilometric radiation. For a review of these types of radio emissions and the associated terminology see ref. 6. RPWS, radio and plasma wave science instrument.

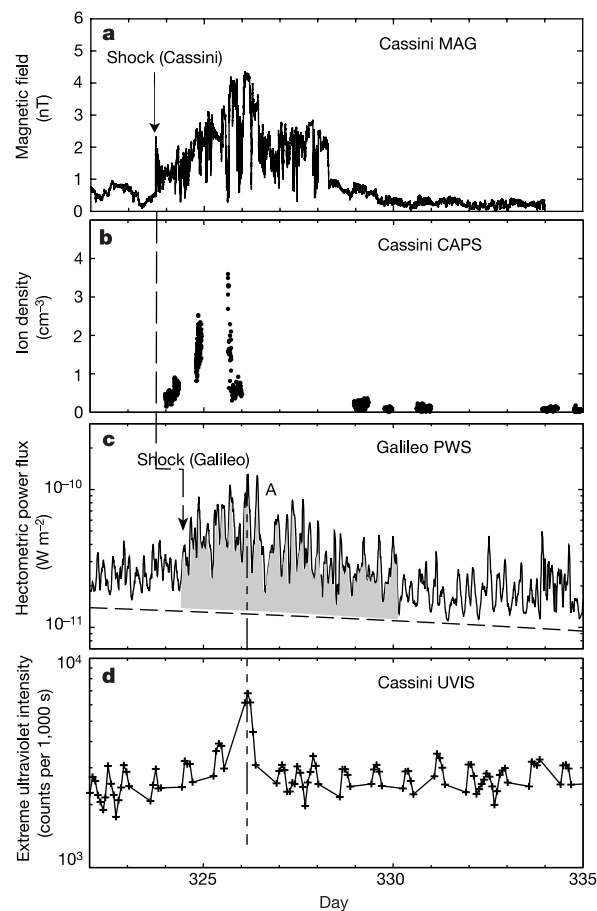




**Figure 2** The power flux of the hectometric radiation as a function of time as detected by the Cassini and Galileo spacecraft during the Cassini fly-by of Jupiter. The power flux is computed by integrating the received spectrum from 0.5 to 5.6 MHz, averaging over the 9 h 55 min rotational period of Jupiter, and correcting to a radial distance of  $100R_J$  using a  $1/R^2$  law. The Cassini plot has been shifted upward by a factor of ten ( $\times 10$ ) relative to the Galileo plot. Three strong hectometric radio emission events (labelled A, B and C) were detected during the Cassini approach to Jupiter. The first starts about midday on day 324 and extends to about day 330, the second starts about day 343 and extends to about day 351, and the third starts late on day 354 and extends to about day 358. Comparisons with the solar wind magnetic field and plasma data from the Cassini magnetometer (MAG)<sup>27</sup> and plasma spectrometer (CAPS)<sup>28</sup> show that event A was preceded by a strong interplanetary shock at 17:92 UT on day 323, event B was preceded by a strong interplanetary shock at 23:12 UT on day 342, and event C was preceded by a strong interplanetary shock at 15:12 UT on day 354. The shocks associated with events A and B were also subsequently detected by the Galileo plasma wave science (PWS) instrument at 09:15 UT on day 324, and 12:10 UT on day 343. The shock associated with event C could not be detected by Galileo, since the spacecraft was well inside the magnetosphere at the time this shock was expected to arrive at Jupiter.

Figure 3c shows that the arrival of the shock at Galileo (indicated by the arrow) is essentially coincident with the onset of the hectometric radio emission event. However, we note that it also takes nearly two days, to early in day 326, before the hectometric radiation reaches peak intensity. When corrected for the solar wind travel time, the peak in the hectometric radiation intensity occurs at about the time of maximum solar wind density. Figure 3d shows that the peak in the hectometric radio emission coincides almost exactly with a peak in the auroral extreme ultraviolet intensity, as detected by the ultraviolet spectrograph on Cassini.

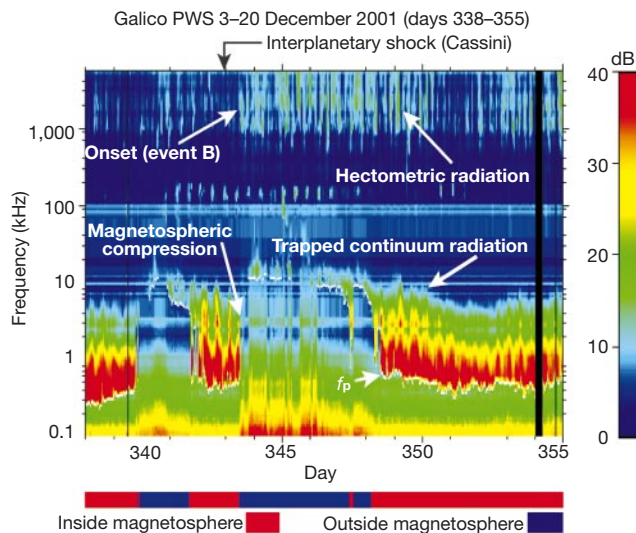
An expanded-timescale frequency–time spectrogram of the radio emission intensities detected by the Galileo spacecraft during event B is shown in Fig. 4. During this period Galileo was very close to the magnetopause on the dusk side of the magnetosphere. The location of the magnetopause relative to the spacecraft can be determined from the presence or absence of trapped continuum radiation, which occurs at frequencies from about 0.5 to 10 kHz. Trapped continuum radiation is present when the spacecraft is inside the magnetosphere, and absent when the spacecraft is outside the magnetosphere<sup>18</sup>. An arrow indicating the time that the interplanetary shock associated with this event was detected by Cassini is shown at the top of the plot. At this time Galileo was clearly within the magnetosphere, as evidenced by the presence of the trapped continuum radiation. At 12:10 UT on day 343, approximately 13 h after the shock was detected by Cassini, the magnetopause abruptly moved inward across Galileo, as indicated by the



**Figure 3** A detailed comparison of the solar wind magnetic field strengths. Data is from the Cassini MAG instrument (a), the solar wind ion densities from the Cassini CAPS instrument (b), the integrated (0.5–5.6 MHz) hectometric radiation intensities (1-h averages) from the Galileo PWS instrument (c), and the disk-integrated extreme ultraviolet auroral H2 band (110.8–113.1 nm) intensities for the Cassini ultraviolet imaging spectrograph (UVIS) instrument<sup>29</sup> (d) for event A. The arrows show the times at which the interplanetary shock was detected by Cassini and Galileo. We note that the onset of the hectometric radiation coincides almost exactly with the arrival of the shock at Galileo. Also the peak in the hectometric radiation intensity occurs at the same time as the maximum brightness of the extreme ultraviolet auroral emission, and near the time of maximum solar wind plasma density (after correcting for the approximately 13-h solar wind travel time).

sudden disappearance of the continuum radiation. Although the shock was not detected directly, this sudden compression was almost certainly caused by the arrival of the interplanetary shock. From the time difference between the arrival of the shock at Cassini (at  $x = +324R_J$ ) and the onset of the compression at Galileo (at  $x = -49R_J$ ) the average propagation speed is estimated to be  $528 \text{ km s}^{-1}$ , which is typical of a low-Mach-number interplanetary shock. As indicated at the top of the spectrogram the onset of the magnetospheric compression corresponds almost exactly with the onset of hectometric radio emission event B.

The observations presented above provide strong evidence that jovian hectometric radiation and auroral extreme ultraviolet emissions are triggered by interplanetary shocks. The processes involved appear to be very similar to those occurring in Earth's magnetosphere. The comparable type of radio emission at Earth is auroral kilometric radiation, which is known to be closely associated with the terrestrial aurora<sup>19</sup>. Just as with the jovian hectometric radiation, the terrestrial kilometric radiation is highly variable and is closely correlated with magnetic storms<sup>20</sup>, including those triggered by interplanetary shocks. Both types of radio emissions are believed to be generated along the auroral magnetic field lines at comparable



**Figure 4** A frequency–time spectrogram from the Galileo PWS showing the radio emission intensities detected by Galileo during event B. The time of the interplanetary shock responsible for this event, as detected by Cassini (at 23:12 UT on day 342), is indicated by the arrow at the top of the spectrogram. The arrival of this shock at Jupiter is believed to be indicated by the disappearance of the trapped continuum radiation at 12:10 UT on day 343, which indicates a sudden compression of the magnetosphere. The white line at the bottom edge of the trapped continuum radiation is the electron plasma frequency, which is proportional to the square root of the electron density. The colour bar at the bottom of the spectrogram indicates when the spacecraft is inside or outside the magnetosphere, on the basis of the presence or absence of the continuum radiation. As can be seen, the onset of the hectometric radiation occurs essentially coincident with the sudden magnetospheric compression on day 343.

altitudes (1 to 3 planetary radii)<sup>14,21</sup> by the cyclotron maser mechanism<sup>10,22</sup>. No *in situ* measurements are available in the high-latitude auroral regions of the jovian magnetosphere, so we use our knowledge of terrestrial auroral processes to help us to understand the processes that are occurring in the jovian magnetosphere. For example, when an interplanetary shock interacts with Earth's magnetosphere, it is known that the magnetosphere becomes strongly compressed. The resulting compression leads to a large-scale reconfiguration of the magnetic field and associated acceleration and energization of plasma within the magnetosphere. The stresses associated with the compression lead to large field-aligned currents and electric fields, particularly along the high-latitude magnetic field lines, where the electrons that carry the field-aligned currents strike the atmosphere and produce the aurora. Very detailed measurements now exist from Earth-orbiting satellites that show how the electrons involved in this process produce the terrestrial auroral kilometric radiation via the cyclotron maser mechanism<sup>23,24</sup>. We infer that similar processes are responsible for the jovian hectometric radiation. At Earth the auroral kilometric radiation is known to be generated along magnetic field lines that terminate in bright spots in the aurora<sup>25</sup>. Because the jovian aurora is also known to have considerable fine structure<sup>26</sup>, we can also infer that the discrete arc-like structure evident in the hectometric radiation (see Fig. 1) is most probably associated with the fine structure in the jovian aurora. □

Received 3 August; accepted 1 November 2001.

- Burke, B. F. & Franklin, K. L. Observations of a variable radio source associated with the planet Jupiter. *J. Geophys. Res.* **60**, 213–217 (1955).
- Drake, F. D. & Hvatum, S. Non-thermal microwave radiation from Jupiter. *Astron. J.* **64**, 329–330 (1959).
- Warwick, J. W. Radiophysics of Jupiter. *Space Sci. Rev.* **6**, 841–891 (1967).
- Broadfoot, A. L. *et al.* Extreme ultraviolet observations from Voyager 1 encounter with Jupiter. *Science* **204**, 979–982 (1979).
- Chamberlain, J. W. *Physics of the Aurora and Airglow* (Academic, New York, 1961).

- Carr, T. D., Desch, M. D. & Alexander, J. K. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 226–284 (Cambridge Univ. Press, Cambridge, 1983).
- Clarke, J. T., Moos, H. W., Atreya, S. K. & Lane, A. L. Observations from Earth orbit and variability of the polar aurora on Jupiter. *Astrophys. J.* **241**, L179–L182 (1980).
- Gurnett, D. A. *et al.* The Cassini radio and plasma wave investigation. *Space Sci. Rev.* (in the press).
- Warwick, J. W. *et al.* Voyager 1 planetary radio astronomy observations near Jupiter. *Science* **204**, 995–998 (1979).
- Wu, C. S. & Lee, L. C. A theory of terrestrial kilometric radiation. *Astrophys. J.* **213**, 621–626 (1979).
- Bigg, E. K. Influence of the satellite Io on Jupiter's decametric radiation. *Nature* **203**, 1008–1010 (1964).
- Goldreich, P. & Lynden-Bell, D. Io, a unipolar inductor. *Astrophys. J.* **156**, 59–78 (1969).
- Queinnee, J. & Zarka, P. Io-controlled decametric arcs and Io–Jupiter interaction. *J. Geophys. Res.* **103**, 26649–26666 (1998).
- Ladreiter, H. P., Zarka, P. & Lecacheux, A. Direction finding study of Jovian hectometric and broadband kilometric radio emissions: Evidence for their auroral origin. *Planet. Space Sci.* **42**, 919–931 (1994).
- Zarka, P. & Genova, F. Low-frequency Jovian emission and solar wind magnetic sector structure. *Nature* **306**, 767–768 (1983).
- Desch, M. D. & Barrow, C. H. Direct evidence for solar wind control of Jupiter's hectometer-wavelength radio emission. *J. Geophys. Res.* **89**, 6819–6823 (1984).
- Gurnett, D. A. *et al.* The Galileo plasma wave investigation. *Space Sci. Rev.* **60**, 341–355 (1992).
- Scarf, F. L., Gurnett, D. A. & Kurth, W. S. Jupiter plasma wave observations: an initial Voyager 1 overview. *Science* **204**, 991–995 (1979).
- Gurnett, D. A. The Earth as a radio source: terrestrial kilometric radiation. *J. Geophys. Res.* **79**, 4227–4238 (1974).
- Voots, G. R., Gurnett, D. A. & Akasofu, S.-Y. Auroral kilometric radiation as an indicator of auroral magnetic disturbances. *J. Geophys. Res.* **82**, 2259–2266 (1977).
- Gallagher, D. L. & Gurnett, D. A. Auroral kilometric radiation: time-averaged source location. *J. Geophys. Res.* **84**, 6501–6509 (1979).
- Zarka, P. Auroral radio emissions at the outer planets: observations and theories. *J. Geophys. Res.* **103**, 20159–20194 (1998).
- Louarn, P. *et al.* Trapped electrons as a free energy source for the auroral kilometric radiation. *J. Geophys. Res.* **95**, 5983–5995 (1990).
- Ergun, R. E. *et al.* Electron-cyclotron maser driven by charge particle acceleration from magnetic field-aligned electric fields. *Astrophys. J.* **538**, 456–466 (2000).
- Huff, R. L. Mapping of auroral kilometric radiation sources to the aurora. *J. Geophys. Res.* **93**, 11445–11454 (1988).
- Clarke, J. T. *et al.* Hubble Space Telescope imaging of Jupiter's UV aurora during the Galileo orbiter mission. *J. Geophys. Res.* **103**, 20217–20236 (1998).
- Dougherty, M. K. *et al.* The Cassini magnetic field investigation. *Space Sci. Rev.* (submitted).
- Young, D. T. *et al.* Cassini plasma spectrometer investigation. *Space Sci. Rev.* (submitted).
- Esposito, L. W. *et al.* The Cassini ultraviolet imaging spectrograph investigation. *Space Sci. Rev.* (in the press).

## Acknowledgements

The research at the University of Iowa, the University of Michigan, and the University of Colorado was supported by NASA through the Jet Propulsion Laboratory.

Correspondence and requests for materials should be addressed to D.A.G. (e-mail: donald-gurnett@iowa.edu).

## Ultra-relativistic electrons in Jupiter's radiation belts

**S. J. Bolton\*, M. Janssen\*, R. Thorne†, S. Levin\*, M. Klein\*, S. Gulkis\*, T. Bastian‡, R. Sault§, C. Elachi\*, M. Hofstadter\*, A. Bunker\*, G. Dulk||, E. Gudim\*, G. Hamilton\*, W. T. K. Johnson\*, Y. Leblanc||, O. Liepack\*, R. McLeod¶, J. Roller¶, L. Roth\* & R. West\***

\* Jet Propulsion Laboratory/Caltech, Pasadena, California 91109, USA

† University of California, Department of Atmospheric Sciences, Los Angeles, California 90024, USA

‡ National Radio Astronomy Observatory, Charlottesville, Virginia 24944, USA

§ Australia Telescope National Facility, Epping, New South Wales 1710, Australia

|| Observatory of Paris, Department of Space Research, F-92195 Meudon, France

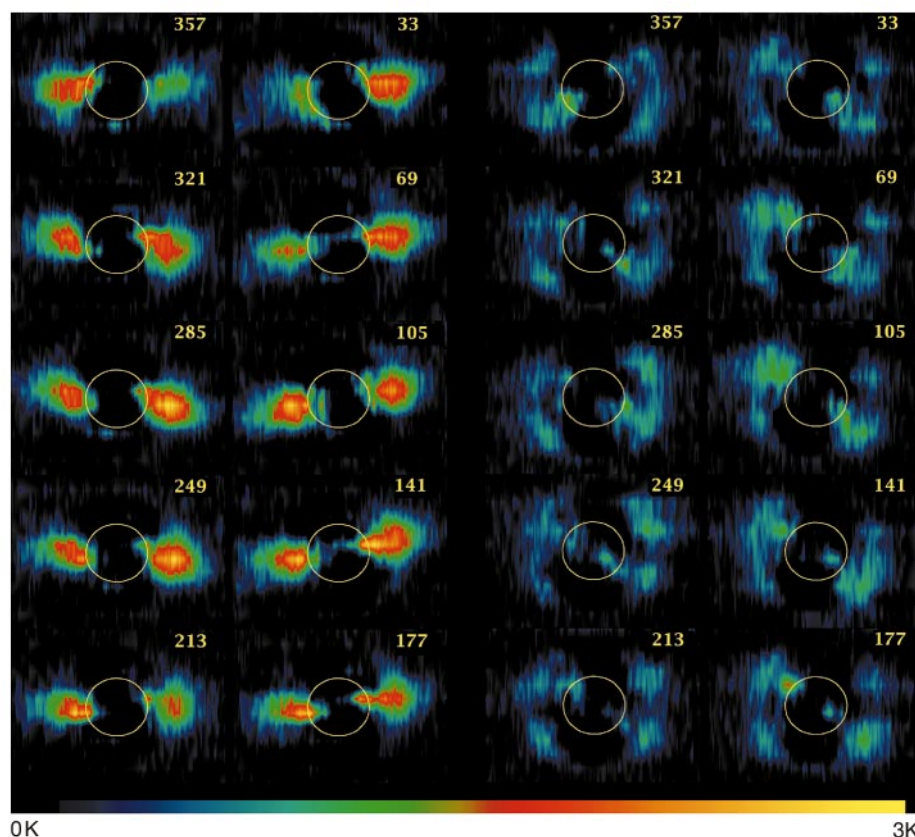
¶ Lewis Center for Educational Research, Apple Valley, California 92307, USA

Ground-based observations have shown that Jupiter is a two-component source of microwave radio emission<sup>1</sup>: thermal atmospheric emission and synchrotron emission<sup>2</sup> from energetic electrons spiralling in Jupiter's magnetic field. Later *in situ* measurements<sup>3,4</sup> confirmed the existence of Jupiter's high-energy electron-radiation belts, with evidence for electrons at

energies up to 20 MeV. Although most radiation belt models predict electrons at higher energies<sup>5,6</sup>, adiabatic diffusion theory can account only for energies up to around 20 MeV. Unambiguous evidence for more energetic electrons is lacking. Here we report observations of 13.8 GHz synchrotron emission that confirm the presence of electrons with energies up to 50 MeV; the data were collected during the Cassini fly-by of Jupiter. These energetic electrons may be repeatedly accelerated through an interaction with plasma waves, which can transfer energy into the electrons. Preliminary comparison of our data with model results suggests that electrons with energies of less than 20 MeV are more numerous than previously believed.

On 2–3 January 2001, *en route* to Saturn, measurements of Jupiter's synchrotron emission were successfully carried out using the radiometer subsystem of the Cassini radar instrument<sup>7</sup> operating at 2.2 cm (13.8 GHz). Synchrotron radiation is the relativistic counterpart to cyclotron radiation (the emission originates from the gyration motion of trapped electrons). Relativistic effects cause the continuum emission to be radiated in a narrow beam focused in

the direction of the electron's motion. The spectrum of synchrotron radiation is characterized by a well defined peak frequency that depends on both electron energy and magnetic field magnitude. Observations at different frequencies (assuming equivalent magnetic field magnitude) relate to emission from electrons at different energies; however, owing to confusion with thermal emission, ground-based telescopes have difficulty observing synchrotron emission at frequencies higher than  $\sim 5$  GHz. Cassini's proximity to Jupiter and the excellent sensitivity of the radar instrument offered an opportunity to map and accurately measure Jupiter's synchrotron emission at a higher frequency than is possible with ground-based telescopes. The motivation was to constrain the high end of the energy spectrum of the electrons trapped in Jupiter's radiation belts. Simultaneous ground-based observations at a wide range of frequencies were carried out to obtain data corresponding to emission from  $\sim 5$  MeV to  $>50$  MeV electrons. Interferometric maps were obtained with the Very Large Array (VLA) operating at 20 cm (1,400 MHz) and 90 cm (333 MHz). Single-dish total flux density measurements were obtained with the Deep Space Network



**Figure 1** Colour maps of Jupiter's synchrotron emission taken by the Cassini spacecraft show the distribution of the ultra-relativistic electrons in Jupiter's inner radiation belts. Maps of horizontal linear polarization (left side) and vertical linear polarization (right side) show bright regions near 1.4 Jupiter radii ( $R_J$ ), corresponding to peak emission by electrons of energy 50 MeV. The observation was carried out shortly after the closest approach to Jupiter, at a distance of  $149R_J$  ( $\sim 10^7$  km) and a phase angle of approximately 75 degrees, using the 4-m on-axis Cassegrain reflector antenna that also serves as the Cassini primary communication antenna. The sub-spacecraft latitude on Jupiter was approximately 1.2 degrees south during the observation. The angular diameter of Jupiter was approximately 0.75 degrees. The colour scale is linear from 0 to 3 kelvin, black to yellow, as shown in the colour bar. Negative values (caused by noise and atmospheric model subtraction) have been set to zero. The 20-h observation was split into two sets of repeating raster scans, obtaining one complete rotation of Jupiter at each of two orthogonal polarizations. Each scan covered approximately  $8R_J$  and  $6R_J$  in east–west and north–south extent, respectively. A single scan, composed of 12 rasters separated by

0.18 degrees or approximately half-beam width (0.23 Jupiter angular diameters), required one hour to complete. Owing to the orientation of the magnetic field, emission originating near the magnetic equator is predominantly linearly polarized in the horizontal sense, while emission originating at higher latitudes is linearly polarized in the vertical sense. To remove thermal emission, a radiative transfer calculation for a nominal model of Jupiter's atmosphere was used to determine the brightness distribution across Jupiter's disk. The antenna beam pattern was determined from a composite of raster scans of the Sun and Jupiter (using an identical technique) obtained before the fly-by. The antenna gain and pointing varied less than 1% and 0.02 mrad, respectively, over the entire 20-h period of the observations. The relative distributions and amplitudes of the residuals as a function of polarization agree well with our model predictions. The variation between maps is related to the beaming curve, and has been studied extensively at 13-cm wavelength by Klein *et al.*<sup>22</sup>. The variation seen in the different panels shows that a form of beaming is present at 2.2 cm.



(DSN) operating at 2.2, 3.5 and 13 cm wavelengths (2.2 and 3.5 cm were used for thermal emission calibration only).

The narrow beaming of the synchrotron emission couples with the non-symmetrical magnetic field and the electron pitch angle distribution to produce variations as Jupiter rotates (the beaming curve). These variations can be seen in the Cassini polarization maps shown in Fig. 1. Cassini unambiguously detected synchrotron emission at a level of  $0.44 \pm 0.15$  Jy (jansky; total integrated flux density adjusted to 4.04 AU). The synchrotron emission was observed to be approximately 1% of the measured thermal emission from Jupiter's disk. The level of synchrotron emission measured was less than estimated from simulations using the energy spectrum and the spatial distribution from current models<sup>5,6</sup>. The results demonstrate the existence of ultra-relativistic electrons near Jupiter and extend the measured nonthermal radio spectrum, providing a new constraint for models of Jupiter's radiation belts.

Observations with the VLA were carried out on 3 January 2001 when Earth was at a jovian declination of  $\sim 3$  degrees. Figure 2 shows the VLA maps at 20 and 90 cm as a function of central meridian longitude. The total synchrotron emission measured was  $5.15 \pm 0.7$  Jy at 90 cm (with this configuration the VLA cannot

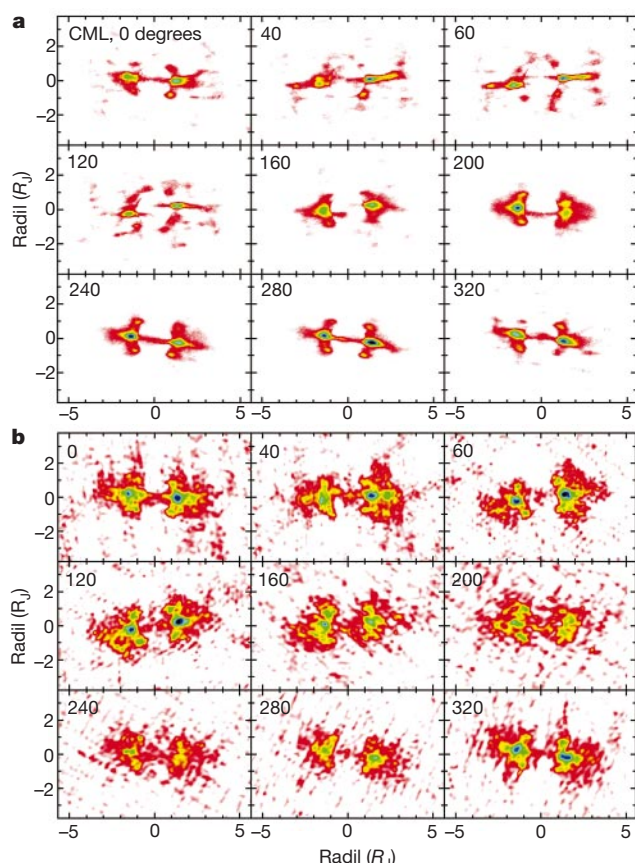
accurately measure the total flux at 20 cm). All four Stokes parameters were measured, although here we present only the images of total intensity, converted to brightness temperature. Observations from the DSN antennas in collaboration with the Goldstone–Apple Valley radio telescope (GAVRT) were obtained from November 2000 to March 2001. The GAVRT project also provided an opportunity for students to work with Cassini scientists, conducting ground-based observations and data analysis. Results from the DSN–GAVRT programme are shown in Fig. 3. The interpolated value of the synchrotron flux density on 3 January was  $4.02 \pm 0.08$  Jy. Short-term variations at the 10% level are evident in the data set.

The new synchrotron emission radio spectrum data is shown in Fig. 4. Data from an earlier epoch at one additional wavelength<sup>8</sup> is shown for completeness. At a frequency of 13.8 GHz (2.2 cm), the region of peak radio emission seen in the maps of Fig. 1 correspond to an average electron energy of  $\sim 50$  MeV gyrating in a background magnetic field of approximately  $\sim 1.2$  G (gauss). The VLA maps at 1.4 GHz and 333 MHz show regions of peak emission corresponding to average electron energies of  $\sim 15$  MeV and  $\sim 7$  MeV, respectively, in a  $\sim 1.2$ -G magnetic field. Because synchrotron radiation is emitted as a continuum, a wide range of electron energies contribute to the maps at each frequency and detailed modelling is required to estimate accurately the electron energy spectrum from the full set of multi-frequency observations. Figure 1 shows significant emission at radial distances  $> 2 R_J$ , suggesting much higher energy electrons ( $\sim 100$  MeV) at the larger radial distances. At a subset of central meridian longitudes, a double radiation belt is possibly indicated with the outer belt located just outside of Jupiter's main ring near  $\sim 1.8 R_J$ , indicative of electron absorption by ring material.

The softer energy spectrum implied by the Cassini high-frequency observations indicates fewer ultra-relativistic electrons ( $\sim 50$  MeV) than predicted by current radiation-belt models. Current models of Jupiter's radiation belts use electron energy spectra based on *in situ* data from spacecraft (Pioneer<sup>3</sup> and Galileo probe<sup>4</sup>) and ground-based synchrotron observations. Although the *in situ* measurements suggested the existence of electrons greater than 20 MeV, the measurements lacked sufficient energy resolution to constrain the energy spectrum at high energies. Simulations of synchrotron emission using the Cassini-derived softer energy spectrum failed to match the intensity level of the lower range of frequencies observed. The preliminary modelling analysis suggests that both a softer energy spectrum and an increase in the number of electrons below 20 MeV are required to match the synchrotron emission spectrum in Fig. 4. The results suggest that the models of Jupiter's radiation belts<sup>5</sup> will probably require more modifications than were previously recommended<sup>9</sup> to estimate accurately the hazards to spacecraft venturing close to Jupiter.

The results reported here have direct effects on our understanding of Jupiter's radiation belts. The detection of synchrotron emission at 13.8 GHz implies that a steady source of  $\sim 50$  MeV electrons must exist. This challenges current theories that have difficulty identifying a source even for the  $\sim 20$ -MeV electrons that are known to exist from previous observations. Secondly, the DSN–GAVRT 13-cm result confirms that the synchrotron emission can vary on relatively short timescales (days), which implies the existence of a process that can rapidly affect the distribution or number of  $\sim 20$ -MeV electrons trapped in Jupiter's strong magnetic field.

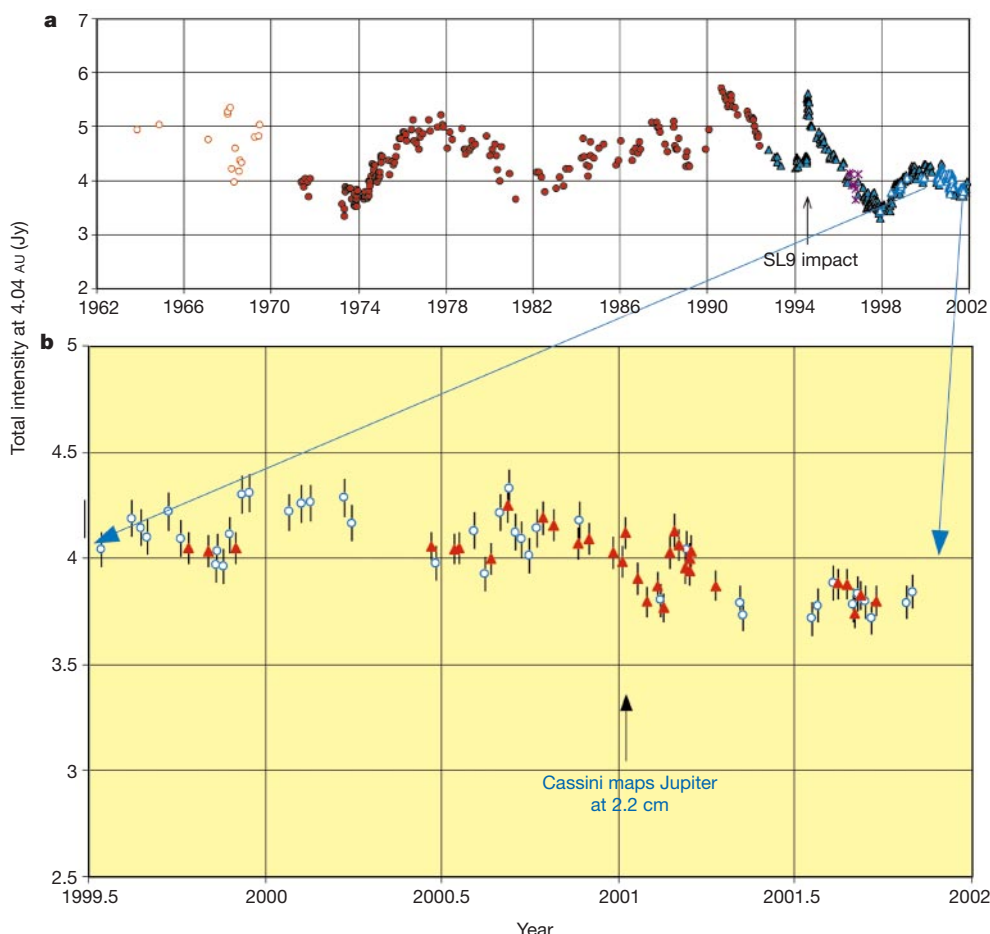
The most widely accepted theory postulates that the relativistic electrons originate in the solar wind or in the outer jovian magnetosphere and diffuse inward, gaining energy through conservation of the first and second adiabatic invariants<sup>10–12</sup>. This can account for  $\sim 10$ -MeV electrons near the peak of the synchrotron zone. However, additional acceleration mechanisms are required to explain higher energies. Efforts to model radial diffusion at Jupiter have failed to reproduce accurately a synchrotron emission radio



**Figure 2** Colour maps of Jupiter's synchrotron radiation taken with the VLA show the distribution of less energetic relativistic electrons (as compared with Fig. 1). The brightest regions near  $1.4 R_J$  correspond to peak emission from electrons of energy 15 MeV (**a**, 20 cm), and 7 MeV (**b**, 90 cm). The observations were made for one full jovian rotation using the VLA operating in A configuration (36 km extent) with a synthesized beam width resolution of approximately 6 arcsec at 90 cm, and 1 arcsec at 20 cm (half-power beamwidth, HPBW). Two-dimensional images were constructed every 40 degrees of central meridian longitude (CML) from data recorded at the given CML  $\pm 40$  degrees. This is a standard construction technique, and thus images from different epochs can be compared directly. After calibration, Jupiter's thermal emission was subtracted (corresponding to a blackbody disk at 350 K), and background confusion sources were removed. For comparison purposes, the VLA data were adjusted to a standard distance of 4.04 AU.

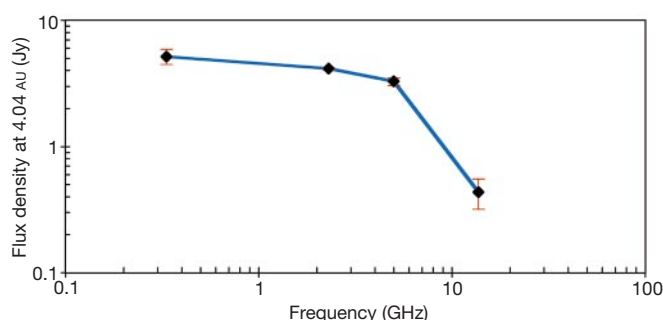
spectrum<sup>12</sup>. Additional sources of electron energization that are known to occur in the Earth's magnetosphere include local stochastic acceleration by wave-particle interactions<sup>13,14</sup>, and acceleration by field-aligned potentials in the auroral zone<sup>15</sup>. Intense wave activity may also be responsible for rapid time variations in

synchrotron emission<sup>16</sup>. Alternative mechanisms, which combine radial diffusive transport with wave-particle interactions, such as recycling and magnetic pumping<sup>17-20</sup>, have also been proposed. Interaction between a very fast interplanetary shock and the Earth's magnetosphere has been proposed as a source of drift resonance



**Figure 3** Variability of Jupiter's synchrotron radio emission shows evidence of long- and short-term variability. **a**, The history of long-term variations in the flux density of Jupiter's synchrotron radio emission at 13-cm wavelength. **b**, An expanded segment of the data taken over the past 2.5 yr that includes the Cassini fly-by period. Thermal flux density from the planet's atmosphere was subtracted from each measurement (assuming a disk temperature of 300 K) and results were normalized to a standard distance of 4.04 AU from Earth. Approximately 2,300 students and their teachers from 26 schools across the United States were part of the ground-based research team working through the GAVRT educational programme. The GAVRT data were merged with the ongoing NASA/JPL

Jupiter Patrol to improve the time resolution of the data (JPL, Jet Propulsion Laboratory). The open circles are NASA/JPL Jupiter Patrol observations made with DSN antennas with apertures of 70 m and 34 m. GAVRT team observations made with 34-m-diameter antennas are represented by filled triangles. The agreement between the two data sets is approximately 1%. These events of 'modest' brightening appear to be intrinsic to Jupiter and are not caused by systematic errors in calibration or by discrete background radio sources that Jupiter passes in front of during its 12-yr orbital path along the ecliptic. Previous observations have suggested the presence of short-term variations in Jupiter's synchrotron emission<sup>23-25</sup>. SL9 indicates comet Shoemaker-Levy 9.



**Figure 4** Jupiter's nonthermal radio spectrum indicates a softer than expected electron energy spectrum at high electron energies. The observed flux density of Jupiter's synchrotron emission observed with Cassini (13.8 GHz) is plotted alongside simultaneous measurements obtained using the DSN (2.3 GHz) and the VLA (0.333 GHz). Previous measurements at 6 cm (5 GHz) are shown for completeness<sup>8</sup>. All measurements are from

January 2001 except for the 6-cm data from 1994. Error bars on the data are indicative of the preliminary nature of the results. Further analysis, including a recent Cassini beam calibration, is expected significantly to reduce the error bars on the data. The ratio between the integrated nonthermal flux densities measured at 2.2 cm and 13 cm is  $S_{2.2}/S_{13} = 0.105 \pm 0.035$ .

acceleration to explain the sudden appearance of relativistic electrons in the terrestrial 'Van Allen' radiation belts<sup>21</sup>. A similar process could exist at Jupiter, although analogies between the terrestrial and jovian magnetosphere are not always appropriate.

Thus, the inner radiation belts of Jupiter and Earth both contain extremely high-energy electrons that require substantial acceleration by processes other than adiabatic radial diffusion. These two magnetospheres are arguably the most explored examples of planetary magnetospheres in the Solar System and in many ways represent the archetype for magnetospheric physics. The fact that both magnetospheres contain electrons that require acceleration processes beyond radial diffusion suggests that this may be a fundamental property of all magnetospheres. This implies that our theoretical understanding of the global properties of magnetospheres may need revision and that the importance of local acceleration mechanisms may currently be underestimated. □

Received 18 September 2001; accepted 17 January 2002.

- Burke, B. F. & Franklin, K. L. Observations of a variable radio source associated with the planet Jupiter. *J. Geophys. Res.* **60**, 213–215 (1955).
- Berge, G. L. & Gulkis, S. in *Jupiter* (ed. Gehrels, T.) 621–692 (Univ. Arizona Press, Tucson, 1976).
- Van Allen, J. A., Baker, D. N., Randall, B. A. & Sentman, D. D. Pioneer 11 observations of energetic particles in the Jovian magnetosphere. *Science* **188**, 459–462 (1975).
- Fischer, H. M., Pehlke, E., Wibberenz, G., Lanzerotti, L. J. & Mihalov, J. D. High-energy charged particles in the inner Jovian magnetosphere. *Science* **272**, 856–858 (1996).
- Divine, N. & Garrett, H. B. Charged particle distribution in Jupiter's magnetosphere. *J. Geophys. Res.* **88**, 6889–6903 (1983).
- Levin, S. M. *et al.* Modeling Jupiter's synchrotron radiation. *Geophys. Res. Lett.* **28**, 903–906 (2001).
- Elachi, C. *et al.* RADAR: The Cassini Titan radar mapper. *Space Sci. Rev.* (in the press).
- Bird, M. K., Funke, O., Neidhofer, J. & De Pater, I. Multi-frequency radio observations of Jupiter at Eufelsberg during the SL9 impact. *Icarus* **121**, 450–456 (1996).
- Bolton, S. J. *et al.* Divine-Garrett model and Jovian synchrotron emission. *Geophys. Res. Lett.* **28**, 907–910 (2001).
- Coroniti, F. V. in *The Magnetospheres of Earth and Jupiter* (ed. Formisano, V.) 391–410 (Reidel, Dordrecht, 1975).
- Bolton, S. J., Gulkis, S., Klein, M. J., de Pater, I. & Thompson, T. J. Correlation studies between solar wind parameters and the decimetric radio emission from Jupiter. *J. Geophys. Res.* **94**, 121–128 (1989).
- de Pater, I. & Goertz, C. K. Radial diffusion models of energetic electrons and Jupiter's synchrotron radiation 1. Steady state solution. *J. Geophys. Res.* **95**, 39–50 (1990).
- Summers, D., Thorne, R. & Xiao, F. Relativistic theory of wave-particle resonant diffusion with application to electron acceleration in the magnetosphere. *J. Geophys. Res.* **103**, 20487–20500 (1998).
- Summers, D., Thorne, R. M. & Xiao, F. Gyroresonant acceleration of electrons in the magnetosphere by superluminal electromagnetic waves. *J. Geophys. Res.* **106**, 10853–10868 (2001).
- Carlson, C. W. *et al.* FAST observations in the downward auroral current region, energetic upgoing electron beams, parallel potential drops, and ion heating. *Geophys. Res. Lett.* **25**, 2017–2021 (1998).
- Bolton, S. J. & Thorne, R. M. Assessment of mechanisms for Jovian synchrotron variability associated with comet SL9. *Geophys. Res. Lett.* **22**, 13–16 (1995).
- Fujimoto, M. & Nishida, A. Monte Carlo simulation of the energization of Jovian trapped electrons by recirculation. *J. Geophys. Res.* **95**, 3841–3854 (1990).
- Nishida, A. Outward diffusion of energetic particles from the Jovian radiation belt. *Geophys. Res. Lett.* **81**, 1771–1773 (1976).
- Sentman, D. D., Van Allen, J. A. & Goertz, C. K. Recirculation of energetic particles in Jupiter's magnetosphere. *Geophys. Res. Lett.* **2**, 465–468 (1975).
- Goertz, C. K. Energization of charged particles in Jupiter's outer magnetosphere. *J. Geophys. Res.* **83**, 3145–3150 (1978).
- Li, X. *et al.* Simulation of the prompt energization and transport of radiation belt particles during the March 24, 1991, SSC. *Geophys. Res. Lett.* **20**, 2423–2426 (1993).
- Klein, M. J., Thompson, T. J. & Bolton, S. J. in *Time Variable Phenomena in the Jovian System* (ed. Bolton, M. J. S.) 151–155 (NASA Spec. Publ. 494, Lowell observatory, 1989).
- Klein, M. J., Gulkis, S. & Bolton, S. J. in *Planetary Radio Emissions IV* (eds Rucker, H. O., Bauer, S. J. & Lecacheux, A.) 217–224 (Austrian Academy of Sciences Press, Vienna, 1996).
- Galopeau, P. H. M., Gerard, E. & Lecacheux, A. Modifications of the synchrotron radiation belts of Jupiter: evidence for natural variations in addition to SL9 effects. *Planet. Space Sci.* **45**, 1197–1202 (1997).
- Miyoshi, Y. *et al.* Observation of short-term variation of Jupiter's synchrotron radiation. *Geophys. Res. Lett.* **26**, 9–13 (1999).

## Acknowledgements

We thank the Cassini Radar team, the Cassini Operations team, C. Hansen and J. Gross. The research reported in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Correspondence and requests for materials should be addressed to S.B. (e-mail: scott.j.bolton@jpl.nasa.gov).

# The dusk flank of Jupiter's magnetosphere

W. S. Kurth\*, D. A. Gurnett\*, G. B. Hospodarsky\*, W. M. Farrell†, A. Roux‡, M. K. Dougherty§, S. P. Joy||, M. G. Kivelson||, R. J. Walker||, F. J. Crayf & C. J. Alexander#

\* Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

† NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

‡ Centre d'étude des Environnements, Terrestre et Planétaires, Université de Versailles Saint-Quentin-en-Yvelines (CETP/UVSQ), 10–12 Avenue de l'Europe, F-78140 Velizy, France

§ Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

|| Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

¶ Space Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan 48109, USA

# Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

Limited single-spacecraft observations of Jupiter's magnetopause have been used to infer that the boundary moves inward or outward in response to variations in the dynamic pressure of the solar wind<sup>1–8</sup>. At Earth, multiple-spacecraft observations have been implemented to understand the physics of how this motion occurs, because they can provide a snapshot of a transient event in progress. Here we present a set of nearly simultaneous two-point measurements of the jovian magnetopause at a time when the jovian magnetopause was in a state of transition from a relatively larger to a relatively smaller size in response to an increase in solar-wind pressure. The response of Jupiter's magnetopause is very similar to that of the Earth, confirming that the understanding built on studies of the Earth's magnetosphere is valid. The data also reveal evidence for a well-developed boundary layer just inside the magnetopause.

The measurements shown here are primarily from the radio and plasma wave science instruments on Cassini<sup>9</sup> and Galileo<sup>10</sup>; they make use of Cassini's fly-by of Jupiter centred on 30 December 2000 coupled with the extended Galileo orbital mission. Figure 1 shows the trajectories of Cassini and Galileo near Jupiter during the time interval surrounding the Cassini closest approach. Cassini first encountered the jovian bow shock at about 0419 spacecraft event time (SCET; HHMM UT at the spacecraft) on 28 December (day 363) 2000. As shown in Fig. 2, the shock is identified in the plasma wave data as a broadband burst of noise extending to about 1.5 kHz. The shock was preceded by about 3 h of Langmuir wave activity<sup>11</sup> at the electron plasma frequency  $f_{pe}$  near 2 kHz, giving an upstream solar-wind electron density  $n_e$  of  $0.05 \text{ cm}^{-3}$  using  $n_e = (f_{pe}/8,980)^2$ , where  $f_{pe}$  is measured in Hz. For several hours preceding the Langmuir wave activity, lower frequency, broadband, ion acoustic wave activity<sup>12</sup> in bursts was observed. We note that the Langmuir waves and ion acoustic waves are mutually exclusive. This would be consistent with Cassini's traversal of the ion foreshock into the electron foreshock. Cassini encountered the bow shock numerous times between 28 December 2000 and the end of the plotted interval. The bow shock was detected as late as early March 2001 to a distance of approximately 800 jovian radii ( $R_J$ ).

On 9 January 2001 between 1250 and 2115 SCET, and again on 10 January between 0655 and 2035 SCET the Cassini radio and plasma wave instrument observed trapped continuum radiation<sup>13,14</sup>, which is a clear indication that the spacecraft was within the jovian magnetosphere. Data from the Cassini plasma and magnetometer instruments confirmed the timing of the magnetopause crossings. A

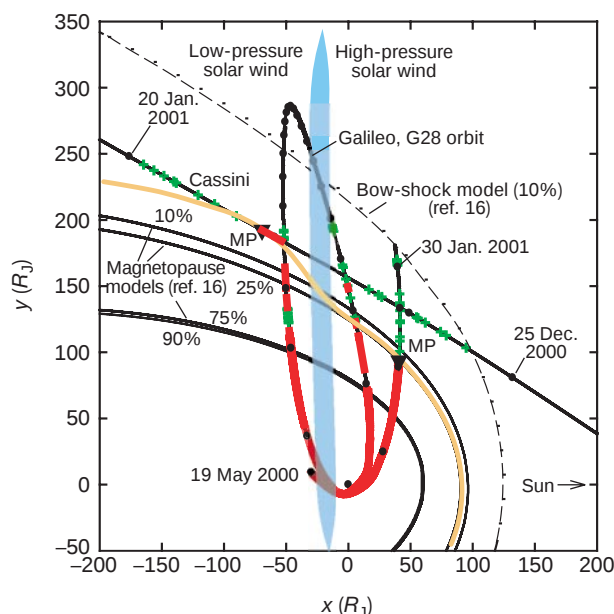


portion of the 10 January wave observations through the magnetopause crossing is shown in the upper panel of Fig. 3. The intense waves between about 300 Hz and 3 kHz are the trapped continuum radiation with the low-frequency cutoff being the electron plasma frequency, corresponding to an electron density of about  $1 \times 10^{-3} \text{ cm}^{-3}$ . The cutoff frequency (electron density) increased to about 2 kHz ( $0.05 \text{ cm}^{-3}$ ) before the continuum radiation disappeared at the magnetopause. The Cassini magnetometer and plasma data confirmed the time of the magnetopause crossing.

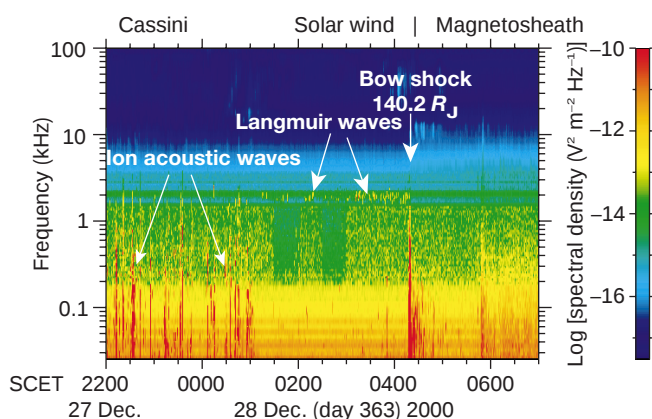
At nearly the same time, Galileo was outbound on its 29th orbit (see Fig. 1) and observed a very similar pattern of continuum radiation as that observed by Cassini, as shown in the bottom panel of Fig. 3. Here, the minimum cutoff frequency of the continuum radiation was somewhat higher, 500 Hz, corresponding to a plasma density of  $3 \times 10^{-3} \text{ cm}^{-3}$ , and the cutoff increased to a frequency of about 3 kHz before disappearing at 2052 SCET at the magnetopause. The time of the magnetopause crossing was confirmed by the Galileo magnetometer. Galileo was well sunward of Cassini; the difference in the  $x$  position of the two spacecraft was more than 100 jovian radii ( $R_J$ ).

The nearly simultaneous magnetopause crossings initially suggest that the magnetopause shape might be approximated by a simple curve connecting the two spacecraft. However, steady-state magne-

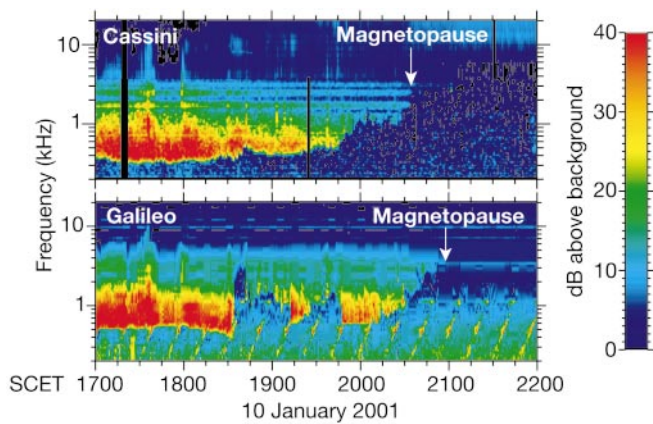
tohydrodynamic models of the interaction of the solar wind with the jovian magnetosphere<sup>15</sup> are inconsistent with this interpretation. Superposed on the trajectories in Fig. 1 are four magnetopause models<sup>16</sup> representing contours on which an observer would have a 10, 25, 75 or 90% chance of being within the magnetosphere, where the 10% contour is farthest from Jupiter. The lower probability curves require smaller solar wind dynamic pressures. Galileo's crossing lies almost directly on the 25% model but the Cassini crossing is significantly more distant than even the 10% (outermost) model<sup>16</sup>. The model magnetopause shapes clearly do not allow a curve representative of a steady-state boundary to intersect both spacecraft. We conclude that the magnetopause was in a state of transition from a significantly inflated size at Cassini, to a large but nominal size at Galileo, and that there must have been a kink or wave in the boundary somewhere between the two as suggested in the qualitative model in Fig. 1. We have assumed that a solar-wind pressure increase is directly transmitted through the magnetosheath at a nominal speed of  $\sim 400 \text{ km s}^{-1}$  and have modelled the reconfiguration of the magnetopause by smoothly connecting the 25% Joy *et al.* contour to one parallel to the Joy *et al.* 10% model<sup>16</sup>, but at a distance consistent with the Cassini crossing. On the basis of magnetohydrodynamic modelling, the radius of curvature of this transition is not extreme. The pressure front would take some five hours to propagate from the position of Galileo to that of Cassini. The Cassini plasma science investigation<sup>26</sup> observed rapidly rising magnetosheath densities over the several hours following this outbound magnetopause crossing, by as much as a factor of ten by 0600 SCET on the following day. Hence, there is clear evidence of a region of increasing pressure moving from Galileo to Cassini's position on timescales similar to what might be expected for a region of increased pressure in the solar wind. Furthermore, after a few short periods of northward magnetosheath fields, Galileo observed continuous southward-directed magnetic fields for approximately a day. Simulations show<sup>17</sup> that southward-directed fields would cause the magnetopause to move outward, so the inward motion observed by both spacecraft could not have resulted from changes of the interplanetary magnetic field orientation. That each spacecraft observed only one outbound crossing on this day rules out a Kelvin–Helmholtz instability, which would have appeared as rather small-amplitude periodic motions of the boundary. Therefore, provided there were no internal variations, we conclude that



**Figure 1** The trajectories of Cassini and Galileo during the time interval surrounding the Cassini closest approach. The coordinate system is centred on Jupiter with the positive  $x$  axis directed from Jupiter to the Sun. The  $z$  axis is normal to Jupiter's orbital plane with positive north. The  $y$  axis completes an orthogonal system. During this time Cassini skimmed the dusk flank of the magnetosphere and for the first three months of 2001 spent a considerable length of time in the dusk jovian magnetosheath, defined as the region between the bow shock and magnetopause. The portion of the Galileo trajectory shown is from its 28th orbit and a portion of the 29th. The apoapses of these orbits were also near dusk, providing some of the first observations of the dusk magnetosphere. On both trajectories, crosses indicate positions where the spacecraft crossed the jovian bow shock. Intervals with thick red lines indicate times when the spacecraft were in the magnetosphere. The triangles (MP) indicate the positions of the two spacecraft when they nearly simultaneously crossed the magnetopause. For comparison, four magnetopause models and one bow-shock model are shown<sup>16</sup>. The bow-shock model illustrated is a surface beyond which there is only a 10% probability of being inside of the shock; that is, it represents the inner boundary of the most distant shocks observed. We notice that Cassini first encountered the bow shock just beyond this model surface. Also included is a qualitative model magnetopause in orange, suggesting that at the time when the two spacecraft crossed the magnetopause roughly simultaneously, the magnetosphere was in transition between states corresponding to two different levels of solar-wind pressure.



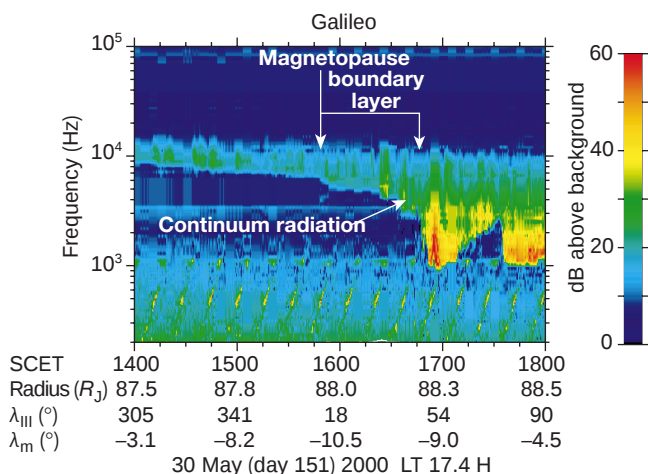
**Figure 2** Plasma wave spectrogram. The initial Cassini bow shock crossing (appearing as the earliest cross on the Cassini trajectory in Fig. 1), upstream ion acoustic waves and Langmuir waves preceding the shock crossing are shown. Here the intensity of waves is shown as a function of frequency (ordinate) and time (abscissa), using the colour bar to indicate electric field spectral density. We note that the Langmuir waves are at a frequency of about 2 kHz, corresponding to an electron density of  $0.05 \text{ cm}^{-3}$ . Assuming a nominal solar-wind speed of  $450 \text{ km s}^{-1}$ , the derived solar wind dynamic pressure is  $0.018 \text{ nPa}$ . The Joy *et al.*<sup>16</sup> 10% bow shock model, which assumes a dynamic pressure of  $0.02 \text{ nPa}$ , crosses the Cassini trajectory very close to this observed shock.



**Figure 3** Simultaneous Cassini and Galileo observations of the jovian magnetopause. In the upper panel, Cassini radio and plasma wave observations show trapped continuum radiation at the beginning of the plotted interval with a low-frequency cutoff of about 300 Hz. The magnetopause crossing time indicated on the spectrogram was determined on the basis of the disappearance of the continuum radiation, changes in the magnetic-field direction and spectrum, and the low-energy plasma, and energetic-particle distributions (D. Mitchell, personal communication). In the bottom panel, similar observations from the Galileo plasma wave instrument are shown. Here the lowest cutoff frequency is somewhat higher than at Cassini.

the magnetosphere was compressed by increasing solar-wind dynamic pressure.

Variations in the size of the terrestrial magnetosphere have been reported for decades, for example, by Fairfield<sup>18</sup>, with Sibeck *et al.*<sup>19</sup> giving a compendium of some 1,821 observations. At Jupiter, fewer, but similar observations have been reported on the basis of Pioneer<sup>1</sup> and Voyager<sup>2–5</sup> data and modelled with conic sections<sup>6–8</sup>. It is well recognized that the magnetopause distance should be controlled, in part, by the solar-wind dynamic pressure<sup>18</sup> and the consequences of these motions have been discussed at length<sup>20–22</sup>. Lepping *et al.*<sup>23</sup> suggested that the extended jovian magnetotail would assume a sausage-like shape owing to recurring pressure variations in the solar wind, suggesting that transients reported herein become a part of the magnetotail structure as they propagate downstream. Kivel-



**Figure 4** Spectrogram showing evidence for a boundary layer just inside the magnetopause. These observations were made on 30 May 2000 by Galileo and show an extended intermediate-density region between the magnetopause as determined by the magnetic field and the low-density magnetospheric lobe. We suggest that this region is analogous to Earth's low-latitude boundary layer.  $\lambda_{III}$ , system III longitude;  $\lambda_m$ , magnetic latitude; LT, local time; H, hours.

son and Southwood<sup>24</sup> argued that in response to changes in pressure at the magnetopause, field-aligned currents would be driven into the auroral ionosphere and suggest that some auroral signatures at high latitudes could result. Although we have found no optical auroral observations concurrent with the observations presented here, Hubble observations<sup>25</sup> taken on 13 January 2001 revealed an auroral oval that was smaller and brighter (upon preliminary analysis) than observed during December 2000 (D. Grodent, personal communication) which might be expected for a compressed magnetosphere.

The gradual increase in electron density inferred from the continuum radiation cutoff from both sets of observations in Fig. 3 is suggestive of a boundary layer just inside the magnetopause where the plasma density changes from that in the magnetosphere to that in the magnetosheath. Galileo showed additional evidence for such a boundary layer during its orbit 28 outbound trajectory. Figure 4 shows a 4-h spectrogram illustrating a rather abrupt decrease in density (decrease in the lower frequency cutoff of the continuum radiation) at about 1645 SCET on 30 May 2000, preceded by a much more gradual decline. The actual magnetopause crossing, determined by a rotation of the field and a decrease in ultra-low-frequency wave activity observed by the magnetometer, occurred at about 1545 SCET. There was a slight decrease in plasma density at this time, but the spacecraft spent nearly an hour in an intermediate density regime, which is also suggestive of a boundary layer. □

Received 1 August; accepted 18 December 2001.

- Smith, E. J., Fillius, R. W. & Wolfe, J. H. Compression of Jupiter's magnetosphere by the solar wind. *J. Geophys. Res.* **83**, 4733–4742 (1978).
- Ness, N. F. *et al.* Magnetic field studies at Jupiter by Voyager 1: Preliminary results. *Science* **204**, 982–987 (1979).
- Ness, N. F. *et al.* Magnetic field studies at Jupiter by Voyager 2: Preliminary results. *Science* **206**, 966–972 (1979).
- Bridge, H. S. *et al.* Plasma observations near Jupiter: Initial results from Voyager 1. *Science* **204**, 987–991 (1979).
- Bridge, H. S. *et al.* Plasma observations near Jupiter: Initial results from Voyager 2. *Science* **206**, 972–976 (1979).
- Lepping, R. P., Burlaga, L. F. & Klein, L. W. Jupiter's magnetopause, bow shock, and 10-hour modulated magnetosheath: Voyagers 1 and 2. *Geophys. Res. Lett.* **8**, 99–102 (1981).
- Acuña, M. H., Behannon, K. H. & Connerney, J. E. P. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 1–50 (Cambridge Univ. Press, Cambridge, 1983).
- Slavin, J. A., Smith, E. J., Spreiter, J. R. & Stahara, S. S. Solar wind flow about the outer planets: Gas dynamic modeling of the Jupiter and Saturn bow shocks. *J. Geophys. Res.* **90**, 6275–6286 (1985).
- Gurnett, D. A. *et al.* The Cassini radio and plasma wave science investigation. *Space Sci. Rev.* (in the press).
- Gurnett, D. A. *et al.* The Galileo plasma wave investigation. *Space Sci. Rev.* **60**, 341–355 (1992).
- Gurnett, D. A., Maggs, J. E., Gallagher, D. L., Kurth, W. S. & Scarf, F. L. Parametric interaction and spatial collapse of beam-driven Langmuir waves in the solar wind. *J. Geophys. Res.* **86**, 8833–8841 (1981).
- Kurth, W. S., Gurnett, D. A. & Scarf, F. L. High resolution spectrograms of ion-acoustic waves in the solar wind. *J. Geophys. Res.* **84**, 3413–3419 (1979).
- Scarf, F. L., Gurnett, D. A. & Kurth, W. S. Jupiter plasma wave observations: An initial Voyager 1 overview. *Science* **204**, 991–995 (1979).
- Gurnett, D. A., Kurth, W. S. & Scarf, F. L. The structure of the Jovian magnetotail from plasma wave observations. *Geophys. Res. Lett.* **7**, 53–56 (1980).
- Ogino, T., Walker, R. J. & Kivelson, M. G. A global magnetohydrodynamic simulation of the Jovian magnetosphere. *J. Geophys. Res.* **103**, 225–235 (1998).
- Joy, S. P., Kivelson, M. G., Walker, R. J., Khurana, K. & Russell, C. T. Probabilistic models of the Jovian magnetopause and bowshock locations. *J. Geophys. Res.* (in the press).
- Walker, R. J., Ogino, T. & Kivelson, M. G. Magnetohydrodynamic simulations of the effects of the solar wind on the Jovian magnetosphere. *Planet. Space Sci.* **49**, 237–245 (2001).
- Fairfield, D. H. Average and unusual locations of the Earth's magnetopause and bow shock. *J. Geophys. Res.* **76**, 6700–6716 (1971).
- Sibeck, D. G., Lopez, R. E. & Roelof, E. C. Solar wind control of the magnetopause shape, location and motion. *J. Geophys. Res.* **96**, 5489–5495 (1991).
- Elphic, R. C. Multipoint observations of the magnetopause: Results from ISEE and AMPTE. *Adv. Space Res.* **8**(9), 223–228 (1988).
- Southwood, D. J. & Kivelson, M. G. The magnetohydrodynamic response of the magnetospheric cavity to changes in solar wind pressure. *J. Geophys. Res.* **95**, 2301–2309 (1990).
- Kivelson, M. G. & Southwood, D. J. Ionospheric traveling vortex generation by solar wind buffeting of the magnetosphere. *J. Geophys. Res.* **96**, 1661–1667 (1991).
- Lepping, R. P. *et al.* Structure and other properties of Jupiter's distant magnetotail. *J. Geophys. Res.* **88**, 8801–8816 (1983).
- Kivelson, M. G. & Southwood, D. J. in *AGU Monograph 58, Physics of Magnetic Flux Ropes* (eds Russell, C. T., Priest, E. R. & Lee, L. C.) 619–625 (AGU, Washington DC, 1990).
- Grodent, D., Clarke, J. T., Kim, J. & Waite, J. H. Jr HST-STIS observations of Jupiter's far-ultraviolet aurora during the Millennium Campaign. *Icarus* (submitted).
- Young, D. T. *et al.* Cassini plasma spectrometer investigation. *Space Sci. Rev.* (in the press).

## Acknowledgements

This research was supported by NASA through contracts with the Jet Propulsion Laboratory.

Correspondence and requests for materials should be addressed to W.S.K. (e-mail: william-kurth@uiowa.edu).

## A nebula of gases from Io surrounding Jupiter

**Stamatios M. Krimigis<sup>\*</sup>, Donald G. Mitchell<sup>\*</sup>, Douglas C. Hamilton<sup>†</sup>, Jannis Dandouras<sup>‡</sup>, Thomas P. Armstrong<sup>§</sup>, Scott J. Bolton<sup>||</sup>, Andrew F. Cheng<sup>\*</sup>, George Gloeckler<sup>†</sup>, K. C. Hsieh<sup>¶</sup>, Edwin P. Keath<sup>\*</sup>, Norbert Krupp<sup>#</sup>, Andreas Lagg<sup>#</sup>, Louis J. Lanzerotti<sup>☆</sup>, Stefano Livi<sup>\*</sup>, Barry H. Mauk<sup>\*</sup>, Richard W. McEntire<sup>\*</sup>, Edmond C. Roelof<sup>\*</sup>, Berend Wilken<sup>#</sup> & Donald J. Williams<sup>\*</sup>**

<sup>\*</sup> Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA

<sup>†</sup> Department of Physics, University of Maryland, College Park, Maryland 20742, USA

<sup>‡</sup> CNES, 9 Avenue du Colonel Roche -BP 4346, F-31028 Toulouse Cedex 4, France

<sup>§</sup> Fundamental Technologies, 2411 Ponderosa, Suite A, Lawrence, Kansas 66046, USA

<sup>||</sup> Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

<sup>¶</sup> Department of Physics, University of Arizona, Tucson, Arizona 85721, USA

<sup>#</sup> Max-Planck-Institut für Aeronomie, D-37191 Katlenburg Lindau, Germany

<sup>☆</sup> Bell Laboratories, 600 Mountain Avenue, Building 1E-439, Murray Hill, New Jersey 07974, USA

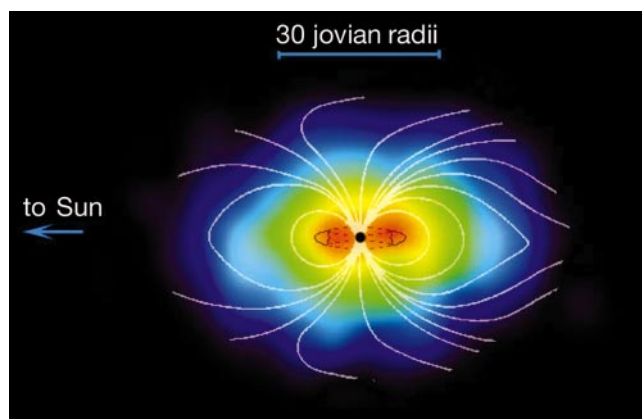
Several planetary missions have reported<sup>1–4</sup> the presence of substantial numbers of energetic ions and electrons surrounding Jupiter; relativistic electrons are observable up to several astronomical units (AU) from the planet. A population of energetic (>30 keV) neutral particles also has been reported<sup>5</sup>, but the instrumentation was not able to determine the mass or charge state of the particles, which were subsequently labelled<sup>6</sup> energetic neutral atoms. Although images showing the presence of the trace element sodium were obtained<sup>7</sup>, the source and identity of the neutral atoms—and their overall significance relative to the loss of charged particles from Jupiter's magnetosphere—were unknown. Here we report the discovery by the Cassini spacecraft of a fast (>10<sup>3</sup> km s<sup>-1</sup>) and hot magnetospheric neutral wind

extending more than 0.5 AU from Jupiter, and the presence of energetic neutral atoms (both hot and cold) that have been accelerated by the electric field in the solar wind. We suggest that these atoms originate in volcanic gases from Io, undergo significant evolution through various electromagnetic interactions, escape Jupiter's magnetosphere and then populate the environment around the planet. Thus a 'nebula' is created that extends outwards over hundreds of jovian radii.

The detectors on the Cassini mission are specifically designed to detect and 'image' such energetic neutral atoms emanating from planetary magnetospheres. The magnetosphere imaging instrument (MIMI)<sup>8</sup> uses an ion and neutral camera (INCA) sensor that rejects charged particles and can detect energetic neutral atoms over the velocity range ~10<sup>3</sup> to ~10<sup>4</sup> km s<sup>-1</sup> using a time-of-flight technique, with crude separation into light (that is, H, He) and heavy (that is, O, S) components. A second MIMI sensor, the charge–energy–mass spectrometer (CHEMS) is capable of determining independently the charge state, mass, and energy of ions over the range of ~3 to ~220 keV per charge.

The presence of Jupiter in terms of energetic neutral atom fluxes became evident soon after the beginning of the Cassini Observatory phase on 1 October 2000, at a distance of ~0.5 AU from Jupiter. The intensity from the peak location in the 32 × 32 pixel image continued to grow, roughly as  $r^{-2}$ , as the spacecraft approached the planet. By the closest approach on 30 December, the image had spread to several pixels. A sample of these images is shown in Fig. 1. There is some evidence of structure in the image, although a significant amount of spreading is due to the point spread function in the instrument (arising from scattering of atoms as they encounter the first foil in the detector)<sup>9</sup>. The intensity is highest at the equatorial plane and extends north and south, as well as in radial distance, roughly as expected from pre-encounter simulations<sup>10</sup>. The velocity distribution in each image pixel is constructed by pulse-height analysis. Although not shown here, it reveals two populations in the measured velocity range corresponding to low and high mass, presumed to be H and O, respectively. Our preliminary view is that Cassini's spectra are consistent with an ion population originating inside about 10 R<sub>J</sub>, where one jovian radius R<sub>J</sub> = 72,400 km.

Detailed measurements of ion composition are obtained by the CHEMS sensor. Figure 2 shows an energy spectrogram of He<sup>+</sup> species for about a month before the first encounter of the jovian bow shock, on day 363. There is significant, low-level activity, above instrument background, throughout this period, interspersed with several intervals of more intense fluxes that are especially evident whenever the spacecraft –x axis was pointed in the general direction of the Sun. Such orientation enables one of three CHEMS sensors



**Figure 1** Energetic neutral atom image of Jupiter's magnetosphere as viewed from the dusk meridian soon after Cassini's closest approach on 30 December 2000. The image was generated by neutrals in the range (3–4) × 10<sup>3</sup> km s<sup>-1</sup> or ~50–80 keV per nucleon,

assuming the species to be hydrogen. The location of Io's plasma torus has been sketched in (dark lines, centre) and Jupiter's magnetic field (white lines) superimposed on the image for reference.

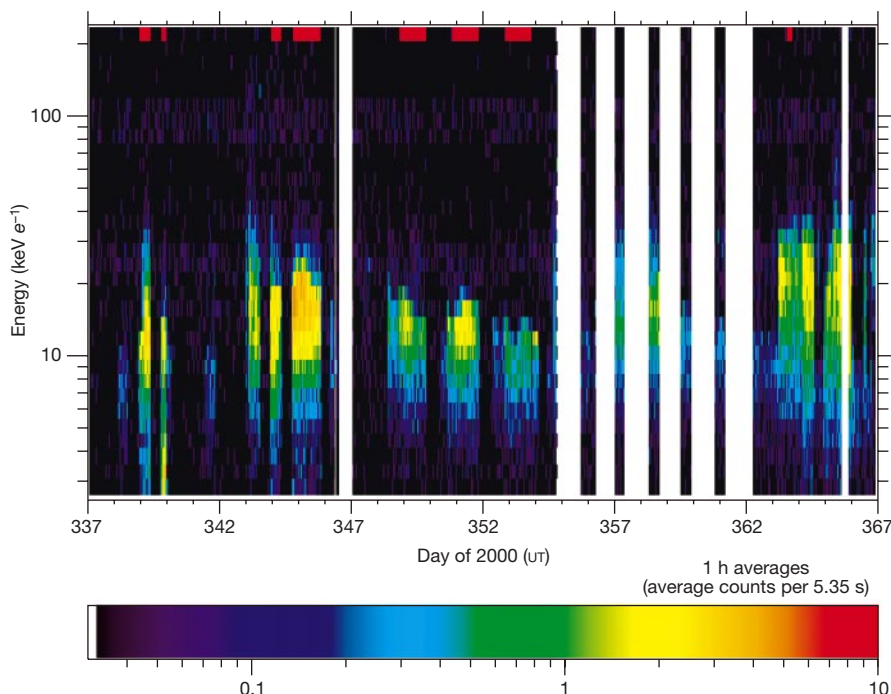


(telescope 2) to view the solar wind direction<sup>8</sup>, and thus become more sensitive to interplanetary pick-up ions<sup>11</sup>. A notable feature is the varying upper energy cutoff in the intensifications ranging from over 20 keV  $e^{-1}$  to  $\leq 10$  keV  $e^{-1}$ ; these cutoffs are directly related to the peak solar wind velocity in the vicinity of the spacecraft, and correspond to  $\sim 500$  km  $s^{-1}$  to  $\sim 350$  km  $s^{-1}$  for high and low cutoffs, respectively. In some cases (for example, on day 343) maximum energies extend much higher, suggesting additional acceleration (see below).

Although  $He^+$  is of interstellar origin<sup>11</sup> and is always present in the interplanetary medium, many ion species observed during this period are of jovian origin. This is shown in Fig. 3, which is a histogram of number of events versus mass per charge summed over all species observed during periods when the Cassini  $-x$  axis is

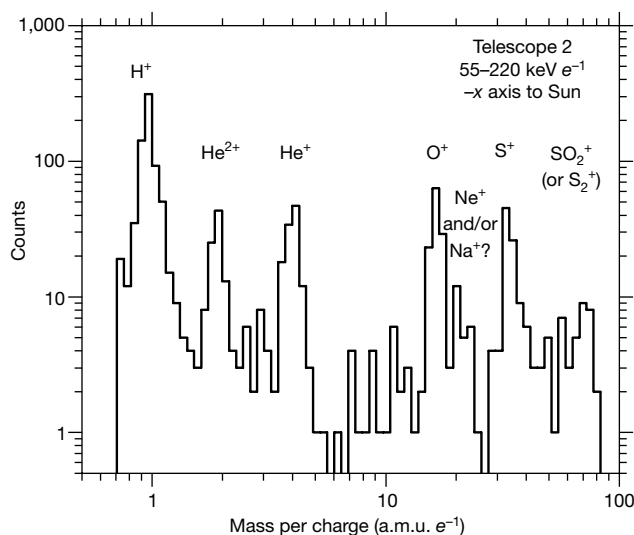
pointing in the solar direction before the first bow shock crossing (days 338–353). In addition to  $H^+$ ,  $He^+$ , and  $He^{2+}$ , there are significant peaks at  $O^+$ ,  $S^+$ , and  $SO_2^+$  (or  $S_2^+$ ). The relatively high energy range for this histogram (55–220 keV  $e^{-1}$ ) emphasizes the heavier pick-up ions of jovian origin, relative to the interstellar  $He^+$ . The  $H^+$  and  $He^{++}$  ions at these energies are not pick-up ions but are instead accelerated by some other process. Statistical uncertainty does not allow identification of other species, although a peak near  $Ne^+$  and  $Na^+$  is suggested.

The results presented above highlight two previously unknown features of the *cis-jovian* environment, shown schematically in Fig. 4. The first is the presence of a continuous flow of fast energetic neutrals in the velocity range  $\sim 10^3$  to  $\sim 10^4$  km  $s^{-1}$ , detectable to

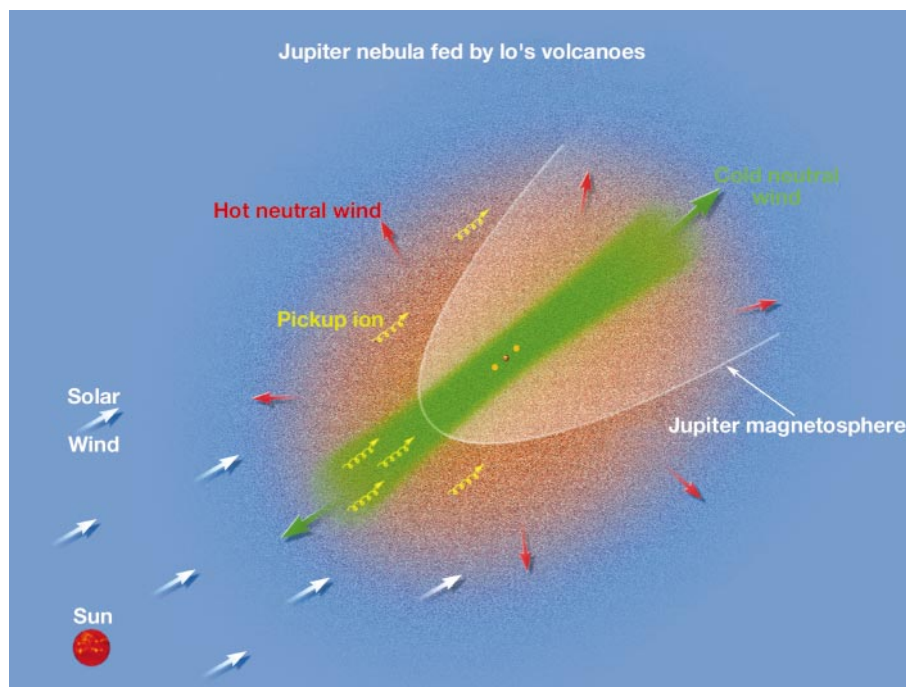


**Figure 2** Energy-time spectrogram of intensity for  $He^+$  ions as measured by the MIMI/CHEMS sensor over the period indicated. The Cassini spacecraft was in the interplanetary medium upstream of Jupiter until day 363, when it encountered Jupiter's bow shock for

the first time. The red bars show periods when the spacecraft  $-x$  axis was pointed in the solar direction and the white spaces correspond to periods of telemetry interruptions.



**Figure 3** Histogram of counts versus mass/charge summed over the days 338–362, 2000, before the Cassini bow shock encounter. Major peaks are identified on the figure. Counts have not been normalized, so no relative elemental abundances are implied.



**Figure 4** Illustrated summary of the findings in this Letter. The boundary of Jupiter's magnetosphere in a meridional cross-section is shown, as is the cross-section of the Io torus. Multicoloured dots are intended to show hot planetary fast neutrals expanding outward in all directions. Green dots represent the cold wind and originate from a lower energy wind of neutral atoms and molecules escaping Io's plasma torus following charge

more than 0.5 AU upstream from the planet. These neutrals probably extend to both lower and higher velocities, and to greater distances from Jupiter than  $\sim 0.5$  AU, beyond the ability of the INCA sensor to measure in its present configuration. The region of energetic neutral atom emission is within  $30 R_J$  of Jupiter (Fig. 1) and appears to be most intense in the vicinity of Io's plasma torus, as we expected.

Assuming spherical symmetry, we estimate the torus loss rate in the measured velocity range to be  $\sim 1-2 \times 10^{26}$  ions  $s^{-1}$ , that is, a small fraction of the overall input rate into the torus of  $\sim 10^{28} s^{-1}$  (ref. 12). Extrapolation to lower velocities can easily increase this rate to  $\sim 10^{27} s^{-1}$ , that is, to a value comparable to the expected loss rate through Jupiter's magnetospheric wind<sup>13</sup>.

The second new feature observed is, by and large, a consequence of the first, in that neutrals escaping Jupiter's magnetosphere, both hot and cold, would be subject to photoionization and thus pick-up by the solar wind electric field<sup>7,14</sup>, attaining a maximum pick-up energy corresponding to twice the solar wind velocity<sup>11</sup>. Furthermore, energetic ions may well escape the magnetosphere and travel upstream in the solar wind under favourable conditions, when the interplanetary magnetic field connects the spacecraft to Jupiter's bow shock<sup>3</sup>. Figure 2 shows examples of both processes. The cutoff energies for  $He^+$  are evidence of the former, while the high energy tails (for example, on days 339, 343 and 354) constitute potential evidence for the latter. Additional study is required to separate escaping energetic ions from those accelerated in interplanetary space. More to the point, the presence of  $O^+$ ,  $S^+$ , and possibly  $SO_2^+$  shows that some fraction of the fast and cold neutral winds from Jupiter becomes reionized and must have some influence in removing energy from the solar wind near Jupiter. Although our analysis of this data set has not progressed far enough to provide a definitive estimate, post-Voyager modelling<sup>12</sup> suggests that this energy reduction could amount to several per cent. This 'jovian nebula' is a permanent element of the *cis*-jovian environment in our Solar System, invisible to distant telescopes for the majority species of O, S,  $SO_2$ , and so on, but visible for the trace element Na (ref. 7). □

exchange and having a corotation speed of  $\sim 75$  km  $s^{-1}$ , generally confined close to Jupiter's equatorial plane. A few neutrals, converted into pick-up ions, are shown gyrating in the general direction of the solar wind, outside the magnetosphere. The solar disk is included for scale.

Received 6 August; accepted 14 November 2001.

1. Teegarden, B. J. *et al.* Interplanetary MeV electrons of Jovian origin. *J. Geophys. Res.* **79**, 3615–3622 (1974).
2. Zwickl, R. D. *et al.* Energetic particle events ( $\geq 30$  keV) of Jovian origin observed by Voyager 1 and 2 in interplanetary space. *J. Geophys. Res.* **86**, 8125–8140 (1981).
3. Krimigis, S. M., Zwickl, R. D. & Baker, D. N. Energetic ions upstream of Jupiter's bow shock. *J. Geophys. Res.* **90**, 3947–3960 (1985).
4. Haggerty, D. & Armstrong, T. P. Observations of Jovian upstream events by Ulysses. *J. Geophys. Res.* **104**, 4629–4642 (1999).
5. Kirsch, E., Krimigis, S. M., Kohl, J. W. & Keath, E. P. Upper limits for X-ray and energetic neutral particle emission from Jupiter: Voyager-1 results. *Geophys. Res. Lett.* **8**, 169–172 (1981).
6. Roelof, E. C., Mitchell, D. G. & Williams, D. J. Energetic neutral atoms ( $E \sim 50$  keV) from the ring current: IMP 7/8 and ISEE-1. *J. Geophys. Res.* **90**, 10991–11008 (1985).
7. Mendillo, M., Baumgardner, J. & Flynn, B. Imaging observations of Jupiter's sodium magneto-nebula during the Ulysses encounter. *Science* **257**, 1510–1512 (1992).
8. Krimigis, S. M. *et al.* Magnetospheric imaging instrument (MIMI) on the Cassini mission to Saturn/Titan. *Space Sci. Rev.* (in the press).
9. Mitchell, D. G. *et al.* The imaging neutral camera (INCA) for the NASA Cassini mission to Saturn and Titan, and possibilities for the future. *Proc SPIE Int. Symp. Opt. Sci. Eng. Instrum.—Mission to the Sun* **2803**, 154–161 (1996).
10. Mauk, B. H., Krimigis, S. M., Mitchell, D. G. & Roelof, E. C. Energetic neutral atom imaging of Jupiter's magnetosphere using the Cassini MIMI instrument. *Adv. Space Res.* **21**(11), 1483–1486 (1998).
11. Gloeckler, G., Geiss, J. & Fisk, L. A. in *The Heliosphere Near Solar Minimum* (eds Balogh, A., Marsden, R. G. & Smith, E. J.) 287–326 (Praxis, Chichester, 2001).
12. Barbosa, D. D. & Eviatar, A. Planetary fast neutral emission and effects in the solar wind: A cometary exosphere analog. *Astrophys. J.* **310**, 927–936 (1986).
13. Krimigis, S. M. *et al.* Characteristics of hot plasma in the Jovian magnetosphere: Results from the Voyager spacecraft. *J. Geophys. Res.* **86**, 8227–8257 (1981).
14. Luhmann, J. G. Oxygen in the heliosphere. *J. Geophys. Res.* **99**, 13285–13305 (1994).

## Acknowledgements

We thank many colleagues who contributed to the hardware effort for MIMI at the Applied Physics Laboratory of Johns Hopkins, the Department of Physics at the University of Maryland, the Max-Planck-Institut für Aeronomie (MPAe) and the Centre d'Etude Spatiale des Rayonnements (CESR) in Toulouse. We also thank S. J. Jaskulek, J. Hayes, M. Kusterer and B. Tossman for their contributions throughout the programme. This work was supported by NASA under contract with the Johns Hopkins University; by subcontracts at the University of Maryland, Kansas, and Arizona; and by Deutsches Zentrum für Luft und Raumfahrt (DLR) at MPAe and Centre National d'Etudes Spatiales (CNES) at CESR.

Correspondence and requests for materials should be addressed to S.M.K. (e-mail: tom.krimigis@jhuapl.edu).

# Ultraviolet emissions from the magnetic footprints of Io, Ganymede and Europa on Jupiter

J. T. Clarke\*, J. Ajello†, G. Ballester‡, L. Ben Jaffel§, J. Connerney||, J.-C. Gérard¶, G. R. Gladstone#, D. Grodent†, W. Pryor\*, J. Trauger† & J. H. Waite Jr‡

\* Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

† Jet Propulsion Laboratory, Pasadena, California 91109, USA

‡ University of Michigan, Ann Arbor, Michigan 48109, USA

§ Institut d'Astrophysique de Paris–CNRS, 98 bis Blvd Arago, Paris 75014, France

|| NASA Goddard Space Flight Center, Greenbelt, Maryland 2077, USA

¶ Université de Liège, 5 Avenue de Cointe, B4000 Liège, Belgium

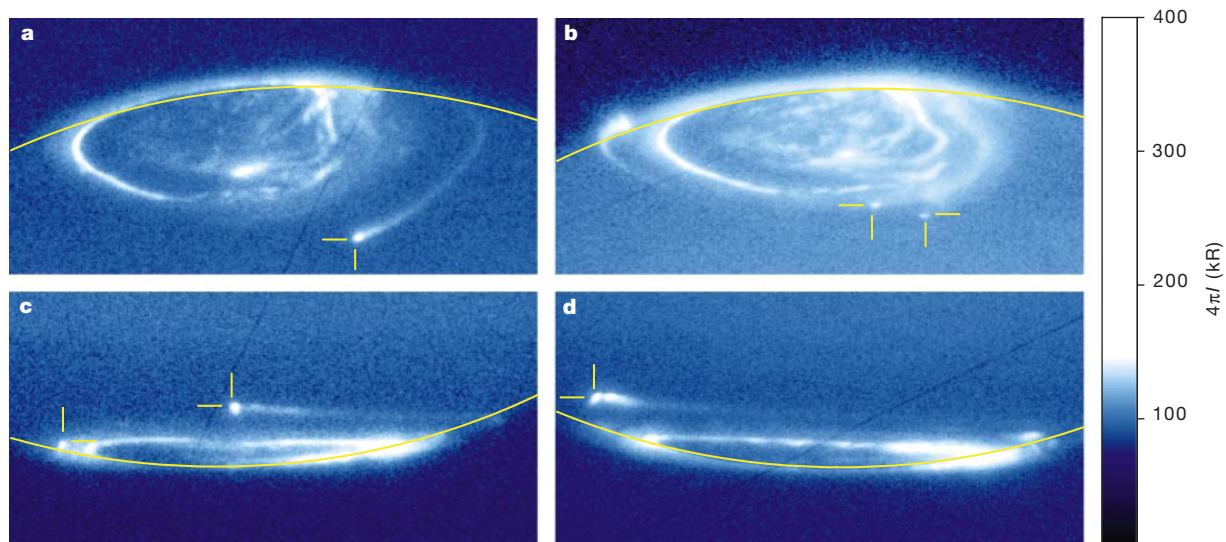
# Southwest Research Institute, Culebra Road, San Antonio, Texas 78228, USA

\* University of Colorado, Boulder, Colorado 80303, USA

Io leaves a magnetic footprint on Jupiter's upper atmosphere that appears as a spot of ultraviolet emission that remains fixed underneath Io as Jupiter rotates<sup>1–3</sup>. The specific physical mechanisms responsible for generating those emissions are not well understood, but in general the spot seems to arise because of an electromagnetic interaction between Jupiter's magnetic field and the plasma surrounding Io, driving currents of around 1 million amperes down through Jupiter's ionosphere<sup>4–6</sup>. The other galilean satellites may also leave footprints, and the presence or absence of such footprints should illuminate the underlying physical mechanism by revealing the strengths of the currents linking the satellites to Jupiter. Here we report persistent, faint, far-ultraviolet emission from the jovian footprints of Ganymede

and Europa. We also show that Io's magnetic footprint extends well beyond the immediate vicinity of Io's flux-tube interaction with Jupiter, and much farther than predicted theoretically<sup>4–6</sup>; the emission persists for several hours downstream. We infer from these data that Ganymede and Europa have persistent interactions with Jupiter's magnetic field despite their thin atmospheres.

Emissions from the magnetic footprints of Io, Ganymede and Europa on Jupiter are presented in Fig. 1, in ultraviolet images taken with the Space Telescope imaging spectrograph (STIS). The main oval, Io footprints, and polar cap emissions have all been observed in earlier Hubble Space Telescope (HST) images<sup>7,8</sup>. Faint emissions have been observed previously scattered around the auroral oval<sup>7</sup>, but the association of these emissions with the satellites can only be established by observing them to remain fixed under the satellites. This was accomplished in a large set of STIS images with high sensitivity and angular resolution obtained between 1998 and 2001. The Ganymede and Europa footprint emissions are indistinguishable from point sources, possibly owing to the limited signal-to-noise ratio. Ganymede's footprint emission appears systematically brighter than Europa's, allowing the measurement of numerous locations of Ganymede's footprint, whereas only a few positions of Europa's have been detected (Fig. 2). Io's footprint emissions at times reach several hundred kilorayleighs (kR) in the measured bandpass ( $1 \text{ kR} = 10^9 \text{ photons cm}^{-2} \text{ s}^{-1} \text{ into } 4\pi \text{ steradians}$ ), corresponding to inputs of tens of  $\text{erg cm}^{-2} \text{ s}^{-1}$  assuming an input electron energy of 30 keV (ref. 9). This represents a large local power input to Jupiter's upper atmosphere and ionosphere. The emissions from the Ganymede and Europa footprints are generally a few tens of kilorayleighs in brightness, corresponding to  $1\text{--}5 \text{ erg cm}^{-2} \text{ s}^{-1}$  or a total power of  $1\text{--}5 \times 10^8 \text{ watts (W)}$ . This compares with up to  $10^{11} \text{ W}$  for the brightest Io footprint emissions, and about half that for the total downstream emissions. Callisto's magnetic footprint overlaps the main oval auroral emissions, so comparably bright emissions would not (and have not) been detected.



**Figure 1** Composite of images with the STIS UV-MAMA (multi-anode microchannel array) of Jupiter's north and south polar regions. **a, c**, 20 January 2001; **b, d**, 26 November 1998. The central meridian longitudes for these images are: **a**, 156.0; **b**, 164.2; **c**, 41.3; and **d**, 14.0 degrees. The extended ultraviolet emissions from Io's magnetic footprint and wake region mapping into Jupiter's ionosphere are the brightest point sources in each image; in **b**, Io's footprint appears above the left-hand limb. The images have been stretched to emphasize the trailing emissions downstream of Io. Ultraviolet emissions from the magnetic footprints of Europa and Ganymede appear in **b** near the central meridian, and Ganymede's footprint appears in **c** near the left-hand limb, indicated with tick marks. Unfiltered images (25MAMA) have a bandpass from 115–174 nm, with sensitivity to the

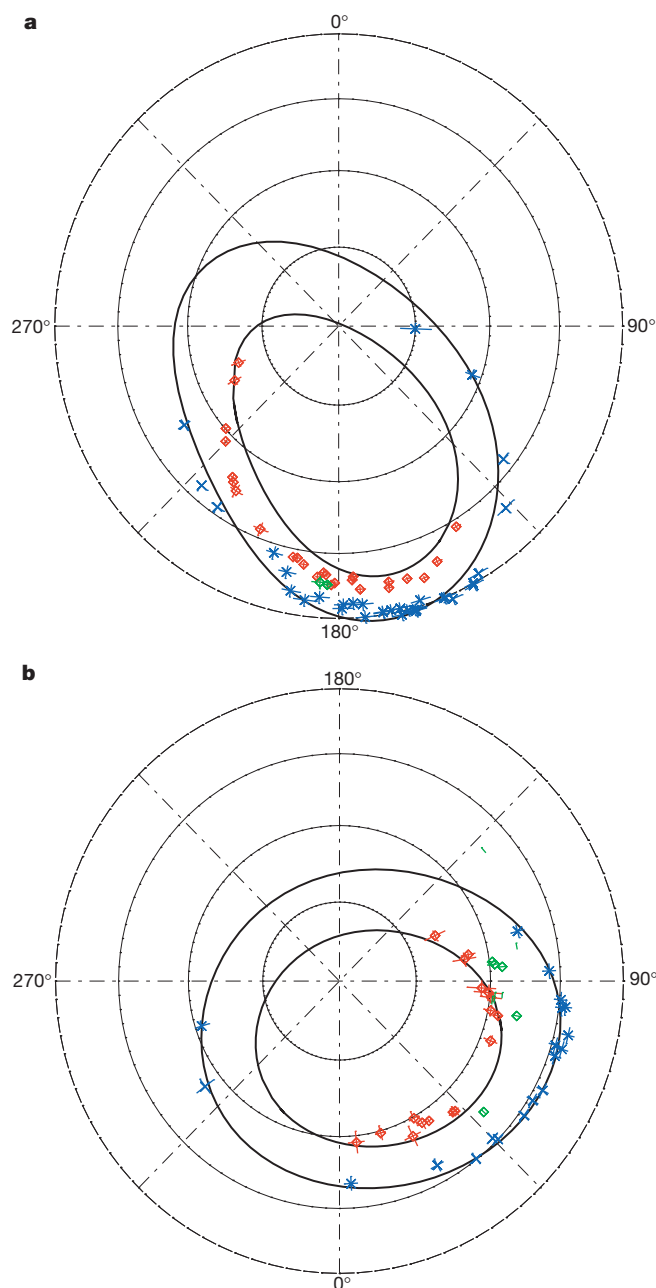
auroral spectrum including the  $\text{H}_2$  Lyman ( $\text{B}^1\Sigma_u^+ \rightarrow \text{A}^1\Sigma_g^+$ ) band emissions and part of the Werner ( $\text{C}^1\Pi_u \rightarrow \text{B}^1\Sigma_g^+$ ) band series plus the H Lyman  $\alpha$  line. All quoted brightnesses and energy deposition rates refer to the flux in this spectral range. The pixel size of 0.024 arcsec provides reasonable sampling of the point spread function of 0.08 arcsec (or 250–300 km on Jupiter); however, the imaging resolution is further degraded by Jupiter's rotation of roughly 1 degree per 100 s. The ability to 'freeze' Jupiter's rotational motion by taking exposures of the order of 100 s with STIS, while still maintaining a limiting detection of 1 kR, provides the highest effective angular resolution and sensitivity obtained on Jupiter's aurora until now, to our knowledge.



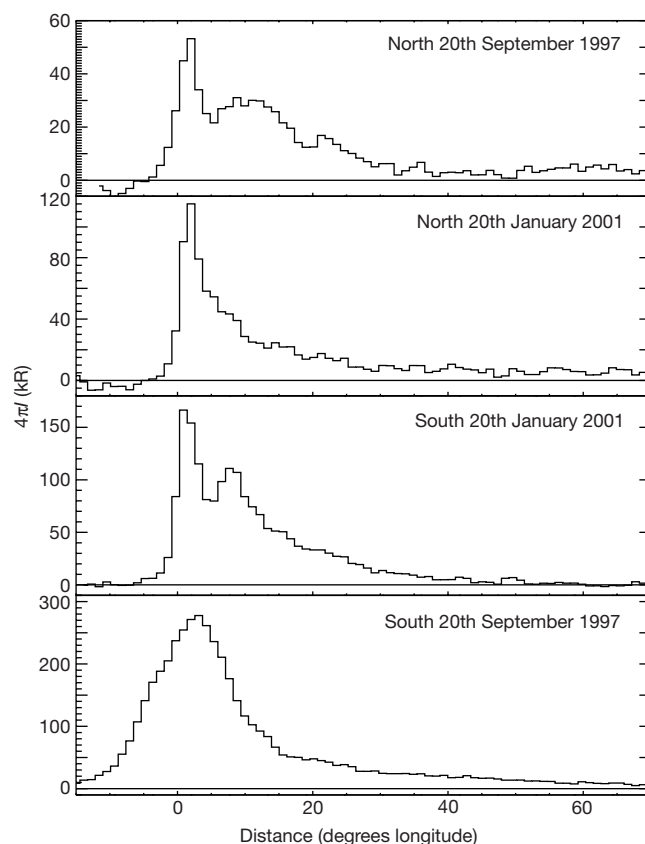
Observations of Ganymede's footprint permit a more accurate determination of the distance to which the main oval maps along Jupiter's magnetic field, and thereby the magnetospheric origin of the auroral activity. Earlier identifications of this distance were based on comparison with planetary magnetic field models, which have a large uncertainty when mapped to the middle magnetosphere owing to the strong radial field component from the current sheet. The identification of Ganymede's magnetic footprint equatorward

of the main oval demonstrates conclusively that the main oval maps to distances greater than Ganymede's 15 Jupiter radius ( $R_J$ ) orbital radius, and the higher latitude of the oval suggests distances greater than  $20R_J$ . The added observation that discrete features in the main oval corotate with Jupiter<sup>7,10</sup> implies a mapping to distances where the plasma nearly corotates, limiting the distance to less than  $30R_J$  because plasma in the magnetodisk begins to fall behind corotation with Jupiter's magnetic field at  $20\text{--}30R_J$ . Recent theoretical papers<sup>11–13</sup> identify this region as the source of the main oval auroral emissions. The observed locations of the Ganymede and Europa footprints will also assist in modelling Jupiter's magnetic field<sup>14</sup>, in the sense that larger latitudinal separations of the loci of satellite footprints from each other and from the main auroral oval (Fig. 2) correspond to weaker field regions.

In addition to Io's footprints, trails of emission are observed extending like a comet's tail in the downstream direction, corresponding to plasma picked up from Io by Jupiter's corotating magnetic field. At a limiting sensitivity of about 1 kR, we have not detected any emissions extending in the upstream direction from Io. Earlier observations had indicated some east–west structure to Io's footprint emissions<sup>7,15</sup>. With an order of magnitude gain in sensitivity over earlier HST cameras, the STIS images show clearly that Io's footprint emission is much more extended than the mapped



**Figure 2** Polar views of north and south auroral zones on Jupiter, showing the observed locations of ultraviolet emissions. The magnetic footprints of Io (blue), Europa (green) and Ganymede (red) are from all STIS images obtained to the end of January 2001. Positional uncertainties are estimated to be 300 and 1,000 km (1 and 3 resolution elements) in the projected distance in the images in the north–south and east–west directions, respectively. **a**, North pole; **b**, South pole. The uncertainties on Jupiter are indicated in this figure by error bars in latitude and longitude. These are generally largest when the emission appears closest to the limb. The overplotted solid lines indicate the mapping from 6 and  $30R_J$  in the VIP4 magnetic field model.



**Figure 3** Brightness traces across the ultraviolet footprint emissions of Io in the direction corresponding to the plasma flow, or approximately east–west on Jupiter. The brightness values are based on an assumed  $0.014 \text{ counts s}^{-1} \text{ kR}^{-1}$  per 0.08 arcsec resolution element for unfiltered images. The spatial extent is large compared with the magnetic mapping of Io's disk onto Jupiter, at times exhibiting a secondary peak in the downstream direction. In the southern aurora, the footprint emission appears double-peaked in the third panel, the left-hand peak corresponding to Io and the other to a point 1,500 km or 10 Io diameters in the downstream direction from Io for an unperturbed field. Owing to the slowing of field lines in Io's wake, however, both emissions may map to regions near Io (see text).  $I$ , intensity into  $4\pi \text{ sr}$ .

projection of Io's disk of about 150 km for the unperturbed magnetic field. This finding, consistent with earlier HST images<sup>7</sup>, resolves the controversy from a report of a strictly local interaction<sup>3,8</sup> based on lower sensitivity measurements.

Within the measurement uncertainty, the downstream trailing emissions overlap the locus of Io footprints. The combined uncertainties of footprint locations and magnetic field modelling are 1–3° in latitude using the VIP4 model<sup>14</sup>, compared with a radial distance of 1R<sub>J</sub> at Io mapping down to about 2° at Jupiter. This limits our knowledge of the region of the plasma torus to which the trail maps, but it is generally along the orbit of Io. The central peaks commonly appear more extended east–west than north–south, with full widths at half maximum in the range of 500 to 3,000 km. In Fig. 1d, the southern footprint displays two peaks of comparable intensity. Brightness scans (Fig. 3) show the distribution of emissions from a sample of images. A secondary emission maximum in the downstream direction is sometimes, but not always, observed. In the northern images it generally appears fainter than the main peak, while the southern emissions appear brighter overall and at times exhibit an equally bright secondary maximum downstream. We have observed separations of the two maxima up to a few thousand kilometres, corresponding to distances of up to 20 Io diameters when mapped along the magnetic field to Io's orbit.

When we interpret these emissions we take into account the Galileo measurement that the local interaction is driven by mass loading, in contrast with the unipolar inductor process<sup>16</sup>. The measured low flow speed of flux tubes in Io's wake<sup>17</sup> implies that they will carry an initial Alfvénic current and lag behind corotation for 1,200 s after first contacting Io<sup>18</sup>. In the downstream region, the flux tubes are brought back up to corotation. The first and last field lines in this 1,200-s interaction with Io correspond to an interval of 12° of Jupiter's rotation, and the observed secondary maxima extend a comparable distance downstream. We may therefore identify the brightest emissions from Io's footprint, including the secondary maximum, as driven by this interaction region near Io. A more detailed modelling of this process will be required to fully understand this interaction.

Io's downstream emissions extend in the plasma flow direction from the instantaneous magnetic mapping of Io for at least 100° in longitude along the magnetic footprint of Io's orbit. This implies active processes that persist for a few hours after Jupiter's magnetic field has swept past Io. We have considered the possibility of a persistent phenomenon, or afterglow, producing ultraviolet emissions after the direct particle precipitation has ended. However, the ultraviolet emissions result from excited upper states in electronic transitions in H and H<sub>2</sub> which are prompt-decaying and produce radiation in a small fraction of a second. One exception is excitation to the 2s state of H, a spin-forbidden transition with a metastable excited state. Ensuing collisions with neutrals can induce radiative decay over a longer timescale; however, a two-photon continuum results from this excited state, limiting the lifetime to seconds<sup>19</sup>. In addition, this process would produce only H Ly $\alpha$  emission, which is inconsistent with the observed bright H<sub>2</sub> spectrum of the downstream emissions. We conclude that the downstream emissions are produced by high-energy charged particles which actively precipitate into Jupiter's atmosphere from the plasma torus downstream of Io, a process continuing at a declining level for several hours after Io has passed.

There are two theoretical frameworks in which the downstream emissions can be interpreted. One is that this emission is produced by plasma stripped from Io but not yet brought up to full corotation. The induction electric field across this plasma could produce a current loop closing in Jupiter's ionosphere, like the interaction at Io itself but with a lower amplitude. The variation in brightness with distance from Io's footprint would then provide an indirect measurement of the rate at which downstream plasma is brought up to

corotation. An alternative interpretation<sup>18,20</sup> involves Alfvén waves near Io propagating along magnetic field lines, which can reflect from the torus boundaries and also from Jupiter's ionosphere. We can imagine numerous geometries of reflecting Alfvén waves propagating downstream, eventually entering the loss cone and producing ultraviolet emissions at Jupiter. This interpretation implies that the downstream ultraviolet emissions could be related to Io's decametric radio arcs<sup>21</sup>. A limitation is that many reflections are needed to extend 100° downstream, and the intensity of any residual waves could decrease to a low level on a shorter spatial scale<sup>20</sup>. We consider the physical process by which the downstream emissions are produced to be uncertain.

To produce ultraviolet emissions from the magnetic footprints of Ganymede and Europa, the initial expectation was for a similar induction process to that occurring at Io but with a smaller amplitude. From the known diameters and orbital distances, plus *in situ* plasma motions from Galileo fly-by measurements, potentials of 100 kV across Ganymede and 150 kV across Europa have been estimated<sup>7</sup>. There is evidence for the modulation of Jupiter's non-thermal radio emissions by Ganymede<sup>22</sup>, and the presence of an intrinsic magnetic field and external magnetosphere on Ganymede<sup>23</sup> presents the interesting possibility of continuous magnetic connection between this satellite and Jupiter. Taking the interaction region to be Ganymede's magnetosphere would increase the cross-sectional area and potential by a factor of four<sup>23</sup>, comparable to the 400-kV potential across Io. Ganymede's interaction clearly leads to footprint emissions which are considerably brighter than Europa's. Localized ultraviolet emissions from Europa and Ganymede themselves have also been detected, indicating that auroral processes are active at these satellites<sup>24,25</sup> at lower intensities than Io's airglow and auroral emissions. Further observations of the satellite and footprint emissions should establish the connections between them. □

Received 13 November 2001; accepted 1 February 2002.

- Connerney, J. E. P. *et al.* Images of excited H<sub>2</sub><sup>+</sup> at the foot of the Io flux tube in Jupiter's atmosphere. *Science* **262**, 1035–1038 (1993).
- Clarke, J. T. *et al.* Far-ultraviolet imaging of Jupiter's aurora and the Io "footprint". *Science* **274**, 404–409 (1996).
- Prangé, R. *et al.* Rapid energy dissipation and variability of the Io–Jupiter electrodynamic circuit. *Nature* **379**, 323–325 (1996).
- Goldreich, P. & Lynden-Bell, D. Io, a Jovian unipolar inductor. *Astrophys. J.* **156**, 59–78 (1969).
- Belcher, J. W. The Jupiter–Io connection, an Alfvén engine in space. *Science* **238**, 170–176 (1987).
- Neubauer, F. M. Nonlinear standing Alfvén wave current system at Io: theory. *J. Geophys. Res.* **85**, 1171–1178 (1980).
- Clarke, J. T. *et al.* Hubble Space Telescope imaging of Jupiter's UV aurora during the Galileo orbiter mission. *J. Geophys. Res.* **103**(E9), 20217–20236 (1998).
- Prangé, R. *et al.* Detailed study of FUV Jovian auroral features with the post COSTAR Hubble Faint Object Camera. *J. Geophys. Res.* **103**, 20195–20210 (1998).
- Dols, V., Gérard, J. C., Clarke, J. T., Gustin, J. & Grodent, D. Diagnostics of the Jovian aurora deduced from ultraviolet spectroscopy: model and GHRS observations. *Icarus* **147**, 251–266 (2000).
- Ballester, G. E. *et al.* Time-resolved observations of Jupiter's far-UV aurora: comparison of WFPC 2 and IUE. *Science* **274**, 409–413 (1996).
- Hill, T. W. The Jovian auroral oval. *J. Geophys. Res.* **106**(A5), 8101–8108 (2001).
- Bunce, E. & Cowley, S. W. H. Divergence of the equatorial current in the dawn sector of Jupiter's magnetosphere: analysis of Pioneer and Voyager magnetic field data. *Planet. Space Sci.* **49**, 1089–1113 (2001).
- Cowley, S. W. H. & Bunce, E. Origin of the main auroral oval in Jupiter's coupled magnetosphere-ionosphere system. *Planet. Space Sci.* **49**, 1067–1088 (2001).
- Connerney, J. E. P., Acuna, M. H., Ness, N. F. & Satoh, T. New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *J. Geophys. Res.* **103**(E9), 11929–11939 (1998).
- Connerney, J., Satoh, T. & Baron, R. H<sub>2</sub><sup>+</sup> emission at the foot of the Io flux tube. *Bull. Am. Astron. Soc.* **27**, 1147 (1996).
- Russell, C. & Huddleston, D. The unipolar inductor myth: mass addition or motional electric field as the source of field-aligned currents at Io. *Adv. Space Res.* **26**, 1665–1670 (2000).
- Frank, L. A. *et al.* Plasma observations at Io with the Galileo spacecraft. *Science* **274**, 394–395 (1996).
- Crary, F. & Bagenal, F. Coupling the plasma interaction at Io to Jupiter. *Geophys. Res. Lett.* **24**, 2135–2138 (1997).
- Bethe, H. & Salpeter, E. *Quantum Mechanics of One and Two Electron Systems* 285 (Plenum, New York, 1977).
- Crary, F. J. On the generation of an electron beam by Io. *J. Geophys. Res.* **102**, 37–49 (1997).
- Wilkinson, M. H. Evidence for periodic modulation of Jupiter's decametric radio emission. *J. Geophys. Res.* **103**(9), 19985–19991 (1998).
- Kurth, W., Gurnett, D. A. & Menietti, J. D. in *Radio Astronomy at Long Wavelengths* 213–217 (AGU Monograph 119, American Geophysical Union, Washington DC, 2000).

23. Kivelson, M. G. *et al.* Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* **384**, 537–541 (1996).
24. Hall, D. T., Feldman, P., McGrath, M. A. & Strobel, D. F. The far-ultraviolet oxygen airglow of Europa and Ganymede. *Astrophys. J.* **499**, 475–481 (1998).
25. Feldman, P. D. *et al.* UV imaging of polar aurora on Ganymede. *Astrophys. J.* **535**, 1085–1090 (2000).

## Acknowledgements

This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA for NASA. The research was supported by grants from the Space Telescope Science Institute and from NASA to the University of Michigan. J.T.C. acknowledges the hospitality of the Institut d'Astrophysique du CNRS in Paris during the preparation of this paper, and J.C.G. and D.G. acknowledge support from the Belgian Fund for Scientific Research and the PRODEX program of ESA. L.B.J. acknowledges support from the Institut National des Sciences de l'Univers and the PNP programme.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T.C. (e-mail: jclarke@bu.edu).

# A pulsating auroral X-ray hot spot on Jupiter

G. R. Gladstone\*, J. H. Waite, Jr†, D. Grodent‡, W. S. Lewis\*, F. J. Crary‡, R. F. Elsner§, M. C. Weisskopf‡, T. Majeed†, J.-M. Jahn\*, A. Bhardwaj§, J. T. Clarke||, D. T. Young†, M. K. Dougherty¶, S. A. Espinosa# & T. E. Cravens\*

\* Southwest Research Institute, San Antonio, Texas 78228, USA

† University of Michigan, Ann Arbor, Michigan 48109, USA

‡ NASA Marshall Space Flight Center, Huntsville, Alabama 35812, USA

§ Vikram Sarabhai Space Centre, Trivandrum, India

|| Boston University, Boston, Massachusetts 02215, USA

¶ Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

# Max-Planck-Institut für Aeronomie, D-37191 Katlenburg-Lindau, Germany

\* University of Kansas, Lawrence, Kansas 66045, USA

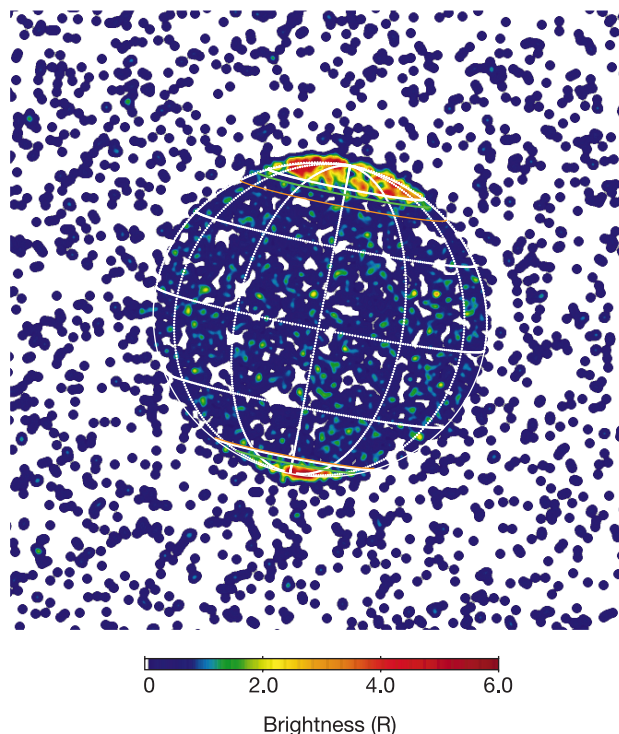
Jupiter's X-ray aurora has been thought to be excited by energetic sulphur and oxygen ions precipitating from the inner magnetosphere into the planet's polar regions<sup>1–3</sup>. Here we report high-spatial-resolution observations that demonstrate that most of Jupiter's northern auroral X-rays come from a 'hot spot' located significantly poleward of the latitudes connected to the inner magnetosphere. The hot spot seems to be fixed in magnetic latitude and longitude and occurs in a region where anomalous infrared<sup>4–7</sup> and ultraviolet<sup>8</sup> emissions have also been observed. We infer from the data that the particles that excite the aurora originate in the outer magnetosphere. The hot spot X-rays pulsate with an approximately 45-min period, a period similar to that reported for high-latitude radio and energetic electron bursts observed by near-Jupiter spacecraft<sup>9,10</sup>. These results invalidate the idea that jovian auroral X-ray emissions are mainly excited by steady precipitation of energetic heavy ions from the inner magnetosphere. Instead, the X-rays seem to result from currently unexplained processes in the outer magnetosphere that produce highly localized and highly variable emissions over an extremely wide range of wavelengths.

Observations were made with the high-resolution camera (HRC) of the Chandra X-ray Observatory on 18 December 2000 for an entire 10-h Jupiter rotation (from 10–20 UT) in support of the Cassini fly-by of Jupiter. These observations show strong auroral emissions from high latitudes (Fig. 1) as well as a rather featureless

disk that probably results from a combination of reflected and fluoresced solar X-rays<sup>11</sup>. The Chandra data are time-tagged and thus can be mapped into jovian latitude and system III longitude coordinates (system III longitudes are based on the 9.925-hour rotation period of Jupiter's magnetic field). Comparison of the resulting X-ray emission map with simultaneous far-ultraviolet images obtained by the Hubble Space Telescope imaging spectrograph (HST-STIS) shows that the northern auroral X-rays are concentrated in a 'hot spot' within the main ultraviolet auroral oval at high magnetic latitudes (Fig. 2).

The hot spot is located roughly at 60–70° north latitude and 160–180° system III longitude; no similar hot spot is seen in the south, but this is almost certainly due to the poor viewing geometry for the southern polar cap. We note that this same hot-spot region is the site of enhanced infrared emissions from CH<sub>4</sub> (ref. 4), C<sub>2</sub>H<sub>2</sub> (ref. 5), C<sub>2</sub>H<sub>4</sub> (ref. 6) and C<sub>2</sub>H<sub>6</sub> (ref. 7), as well as highly variable H<sub>2</sub> emissions at far-ultraviolet wavelengths<sup>8</sup>, and a 'dark spot' in the sunlight reflected from Jupiter at mid-ultraviolet wavelengths<sup>12</sup>.

Jupiter's main auroral oval lies at latitudes that map magnetically to radial distances near 30 jovian radii, R<sub>J</sub> (refs 13–15); the location of the hot spot at latitudes poleward of the main oval indicates that the bulk of the jovian X-ray emissions must connect along magnetic field lines to regions in the jovian magnetosphere well in excess of 30R<sub>J</sub> from the planet. The Chandra HRC observations therefore call into question earlier views that attribute the X-ray auroral emissions to energetic particles diffusing planetward from the outer regions of the Io plasma torus and precipitating in the atmosphere at latitudes



**Figure 1** Chandra X-ray Observatory image of Jupiter on 18 December 2000. False colour brightnesses are indicated in rayleighs (R). The observation lasted 10 h (10–20 UT) and each X-ray photon has been smeared by double the 0.4-arcsecond full-width half-maximum point-spread-function of the high-resolution camera. A jovian-centric graticule with 30° intervals is overplotted, along with the maximum equatorward extent of the  $L = 5.9$  (orange lines) and  $L = 30$  (green lines) footprints of the VIP4 model<sup>16</sup> magnetosphere. The auroral emissions are located at much higher latitudes than we expected on the basis of previous X-ray observations and indicate a connection with Jupiter's outer magnetosphere. An animation showing the time dependence of these observations may be viewed at [http://pluto.space.swri.edu/yosemite/jupiter/chandra\\_hrc.html](http://pluto.space.swri.edu/yosemite/jupiter/chandra_hrc.html).



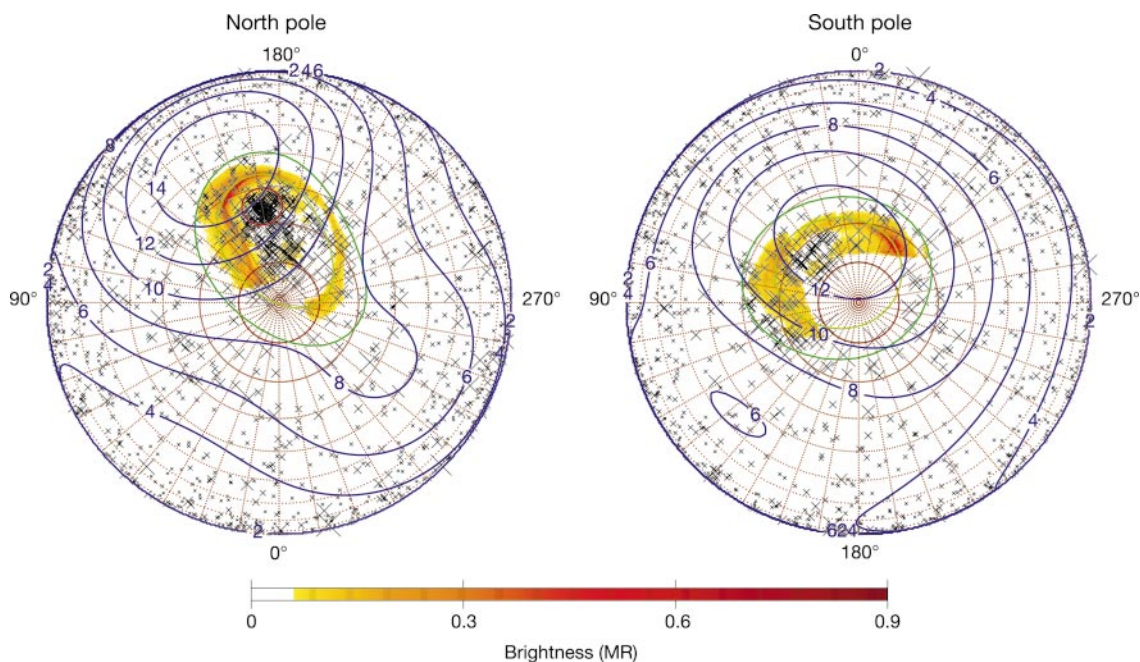
below those of the main oval<sup>3</sup>. On the other hand, our magnetic mapping of the hot spot to distances greater than  $30R_J$  means that the source of the precipitating particles is unclear, because at such large distances from Jupiter there are insufficient S and O ions (B. H. Mauk, personal communication) to account for the hot-spot emissions. Another ion source or excitation mechanism (such as electron bremsstrahlung) must be considered.

Further evidence that some process other than energetic ion precipitation from the inner magnetosphere is responsible for the bulk of the observed auroral X-rays is provided by the lack of expected correlation between the X-ray emission morphology and the surface magnetic field strength (that is, the magnetic field strength at  $1R_J$ ) as determined with the VIP4 model<sup>16</sup> (Fig. 2). That is, for the nominal mechanism of generation by energetic ion precipitation, the brightest X-ray emissions would be expected where the eastward drifting (that is, toward lower longitude) ions encounter the most steeply decreasing surface magnetic field strength along their L-shell footprint (that is, the locus of intersection of their magnetic field lines with the surface of Jupiter) and only if the field strength is lower than in the conjugate hemisphere<sup>17,18</sup>. Thus, although we would expect emissions at slightly higher latitudes than the  $L = 5.9$  footprint of the Io plasma torus, at system III longitudes of  $0$ – $60^\circ$  in the north and  $120$ – $260^\circ$  in the south, we found minor clusters of X-rays near the  $L = 5.9$  footprint near  $140^\circ$  in the north and  $80$ – $120^\circ$  in the south.

A result even more puzzling than the high-latitude location of the X-ray hot spot is revealed when the X-ray counts are plotted as a function of time. The resulting light curve and power spectrum (Fig. 3) show a very strong  $\sim 45$ -min oscillation in the emitted X-rays. One of the primary goals of the Chandra and HST campaigns supporting the Cassini fly-by was to search for transient auroral variations that might be related to the interaction of the solar wind

with Jupiter's magnetosphere. However, correlative Cassini solar-wind data acquired upstream at about  $200R_J$  show no comparable periodicity, even accounting for the 5–10-h delay time for the propagation from the spacecraft to the planet. Likewise, no 45-min periodicities were seen in Galileo and Cassini energetic-particle and plasma-wave measurements at the time of the Chandra observations, although such periodicities are seen at other times (W. S. Kurth, personal communication). Forty-minute oscillations have been seen before in energetic particles in the outer magnetosphere and in radio waves<sup>9,10</sup>. Following the Ulysses fly-by of Jupiter, intermittent bursts of 1–200-kHz radio emissions with an approximately 40-min period were observed for several months originating from high southern-jovian latitudes; these bursts were correlated with Ulysses measurements of solar-wind velocity and both relativistic ( $>8$  MeV) and lower ( $\sim 50$  keV) energy electrons from Jupiter<sup>9</sup>. However, the origin of these quasi-periodic radio bursts has not been explained.

As there is no apparent correlation between the auroral X-rays and the solar-wind parameters measured by Cassini before and during the Chandra observations, it seems most likely that the oscillations arise from processes internal to the jovian magnetosphere. Global ultra-low-frequency (ULF) oscillations of the magnetic field and of the density of high-energy ions are ubiquitous in the jovian magnetosphere and are generally found to have periods in the 10–20-min range<sup>19,20</sup>. Certain models of the ULF oscillations as standing waves along magnetic field lines indicate that spacecraft motion affects the measured periods so that they are closer to one hour in a reference frame that corotates with Jupiter<sup>19</sup>. The observed ULF oscillations may arise in a resonance with the bounce periods of the energetic particles (that is, the period for a magnetically trapped ion to repeat its north–south motion along a field line). Scattering of some portion of this particle population into the loss cone could



**Figure 2** Polar projections of X-rays seen by Chandra and simultaneous far-ultraviolet images obtained by the Hubble Space Telescope. The mapped locations of individual X-ray photons (crosses) are overlaid on averages of several northern (left) and southern (right) auroral images made with the Hubble Space Telescope imaging spectrograph (HST-STIS) during 10–20 UT on 18 December 2000. The mapping assumes that the X-ray and ultraviolet auroras peak in emission at 240 km above the 1-bar pressure level. The size of each cross gives an approximate indication of the uncertainty in location of the corresponding X-ray photon, and only photons with emission angles of  $<85^\circ$  are shown. The HST-STIS images made with the 25MAMA filter are displayed in false colour with

auroral  $H_2$  emission brightnesses in megarayleighs (MR) as indicated by the colour bar. Surface VIP4 model<sup>16</sup> magnetic field strength contours are shown for comparison (dark blue). The  $L = 5.9$  and  $L = 30$  footprints of the VIP4 model magnetosphere are also included (outer and inner green ovals, respectively), and a  $10^\circ$  graticule (brown dotted lines) with system III longitudes labelled. Most of the northern auroral X-rays are unexpectedly located well within the main far-ultraviolet oval and are coincident with the polar-cap far-ultraviolet emissions. The red circle in the northern auroral plot (left) shows the region defined for the hot spot used in the timing analysis. The apparent increase in X-rays toward the equator is an artefact of the polar projection.

result in quasi-periodic precipitation that would account for the periodicity observed in the X-ray emissions. However, bounce periods vary with particle energy, distance from Jupiter, and pitch angle, and it is unclear what would cause a narrow range of periods to dominate this resonance over much of the magnetosphere.

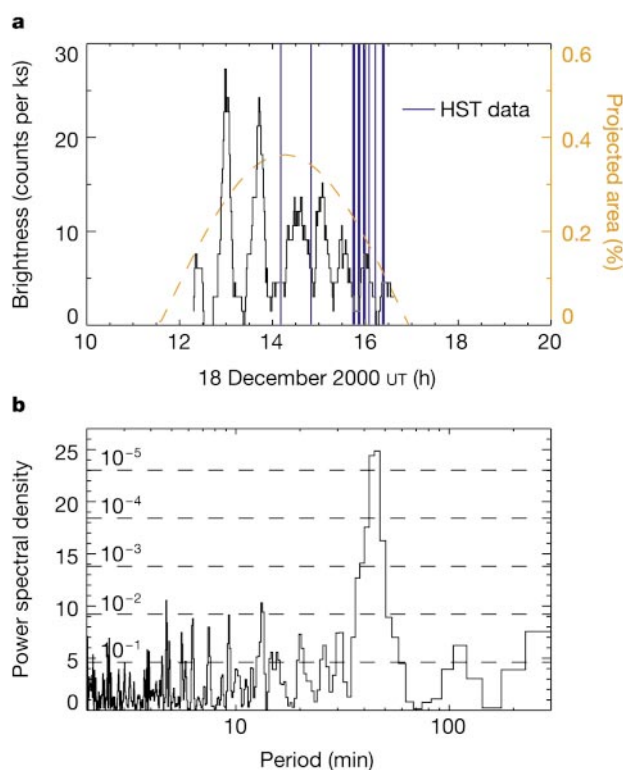
It is difficult to estimate the power in the emitted X-rays, because the Chandra HRC responds over a broad energy range (0.1–10 keV) with a variable sensitivity that peaks near an energy of 1.1 keV. We currently have no knowledge of the details of the emitted spectrum, so we can only make very rough estimates of the emitted power. Assuming a photon energy of 574 eV (corresponding to an  $O^{6+}$  emission feature expected to be bright in ion auroras or solar-wind charge exchange<sup>21,22</sup>), the estimated X-ray luminosities of the disk of Jupiter and its northern and southern auroras are about 2.3, 1.0 and 0.4 GW, respectively. These results are consistent with previous observations made with low spatial resolution<sup>1,2</sup>.

As we note above, it is difficult to account for the ion flux needed to produce the estimated luminosities with a source region located in the outer magnetosphere. If the emissions are indeed generated by heavy-ion precipitation, one possibility is high-latitude recon-

nection of the planetary and solar-wind magnetic fields, with the subsequent entry of the highly ionized (but low energy) heavy-ion component of the solar wind. The captured solar-wind ions could be accelerated to MeV energies by the field-aligned currents present in the outer magnetosphere<sup>22–24</sup>. Such particles could also be consistent with the observed plasma waves. For example, the bounce period of 20 MeV oxygen ions on a dipole field line at  $L = 120R_J$  with an equatorial pitch angle of  $30^\circ$  is about 38 minutes. Although outer magnetospheric field lines are not dipolar<sup>25</sup>, they are close enough for this simple calculation to be informative. We wondered whether electron bremsstrahlung, originally rejected primarily on energetic grounds, should be reconsidered as an explanation for the X-rays. The energetics argument still holds: the power needed to produce the brightest far-ultraviolet ‘flares’ seen in the same polar-cap region as the X-ray hot spot is a few tens of TW (ref. 8), much less than the estimated power of a few PW (ref. 1) needed to produce the observed X-rays by electron bremsstrahlung. Thus, explaining the observed hot-spot X-rays with electron bremsstrahlung still seems unpromising. Whatever ultimate source is determined for the hot-spot X-rays, it should probably also account for the far-ultraviolet flare emissions, the various hydrocarbon infrared emissions, and possibly the mid-ultraviolet dark spot, as it is unlikely that these various phenomena occur in the same area of the upper atmosphere of Jupiter and yet are unrelated to one another. □

Received 3 August; accepted 18 December 2001.

1. Metzger, A. E. *et al.* The detection of x-rays from Jupiter. *J. Geophys. Res.* **88**, 7731–7741 (1983).
2. Bhardwaj, A. & Gladstone, G. R. Auroral emissions of the giant planets. *Rev. Geophys.* **38**, 295–353 (2000).
3. Mauk, B. H. *et al.* Hot plasma parameters of Jupiter's inner magnetosphere. *J. Geophys. Res.* **101**, 7685–7695 (1996).
4. Caldwell, J. J., Halthore, R., Orton, G. & Bergstralh, J. Infrared polar brightenings on Jupiter. 4. Spatial properties of methane emission. *Icarus* **74**, 331–339 (1988).
5. Drossart, P. *et al.* Enhanced acetylene emission near the north pole of Jupiter. *Icarus* **66**, 610–618 (1986).
6. Kostiuik, T., Romani, P., Espenak, F., Livengood, T. A. & Goldstein, J. J. Temperature and abundances in the Jovian auroral stratosphere. 2. Ethylene as a probe of the microbar region. *J. Geophys. Res.* **98**, 18823–18830 (1993).
7. Livengood, T. A., Kostiuik, T., Espenak, F. & Goldstein, J. J. Temperature and abundances in the Jovian auroral stratosphere. 1. Ethane as a probe of the millibar region. *J. Geophys. Res.* **98**, 18813–18822 (1993).
8. Waite, J. H. Jr *et al.* An auroral flare at Jupiter. *Nature* **410**, 787–789 (2001).
9. MacDowall, R. J. *et al.* Quasiperiodic jovian radio bursts: Observations from the Ulysses radio and plasma wave experiment. *Planet. Space Sci.* **41**, 1059–1072 (1993).
10. Krimigis, S. M. *et al.* A nebula of gases from Io surrounding Jupiter. *Nature* **415**, 994–996 (2002).
11. Maurellis, A. N. *et al.* Jovian x-ray emission from solar x-ray scattering. *Geophys. Res. Lett.* **27**, 1339–1342 (2000).
12. Vincent, M. B. *et al.* Jupiter's polar regions in the ultraviolet as imaged by HST/WFPC2: Auroral-aligned features and zonal motions. *Icarus* **143**, 205–222 (2000).
13. Connerney, J. E. P., Baron, R., Satoh, T. & Owen, T. Images of excited  $H_2^+$  at the foot of the Io flux tube in Jupiter's atmosphere. *Science* **262**, 1035–1038 (1993).
14. Clarke, J. T. *et al.* Hubble Space Telescope imaging of Jupiter's UV aurora during the Galileo orbiter mission. *J. Geophys. Res.* **103**, 20217–20236 (1998).
15. Prangé, R. *et al.* Detailed study of FUV jovian auroral features with the post-COSTAR HST faint object camera. *J. Geophys. Res.* **103**, 20195–20215 (1998).
16. Connerney, J. E. P., Acuña, M. H., Ness, N. F. & Satoh, T. New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *J. Geophys. Res.* **103**, 11929–11939 (1998).
17. Gehrels, N. & Stone, E. C. Energetic oxygen and sulfur ions in the jovian magnetosphere and their contribution to the auroral excitation. *J. Geophys. Res.* **88**, 5537–5550 (1983).
18. Herbert, F., Sandel, B. R. & Broadfoot, A. L. Observations of the jovian UV aurora by Voyager. *J. Geophys. Res.* **92**, 3141–3154 (1987).
19. Khurana, K. K. & Kivelson, M. G. Ultralow frequency MHD waves in Jupiter's middle magnetosphere. *J. Geophys. Res.* **94**, 5241–5254 (1989).
20. Wilson, R. J. & Dougherty, M. K. Evidence provided by Galileo of ultra low frequency waves within Jupiter's middle magnetosphere. *Geophys. Res. Lett.* **27**, 835–838 (2000).
21. Liu, W. & Schultz, D. R. Jovian x-ray aurora and energetic oxygen ion precipitation. *Astrophys. J.* **526**, 538–543 (1999).
22. Cravens, T. E. Comet Hyakutake x-ray source: Charge transfer of heavy solar wind ions. *Geophys. Res. Lett.* **24**, 105–108 (1997).
23. Bunce, E. J. & Cowley, S. W. H. Local time asymmetry of the equatorial current sheet in Jupiter's magnetosphere. *Planet. Space Sci.* **49**, 261–274 (2001).
24. Krupp, N. *et al.* Local time asymmetry of energetic ion anisotropies in the jovian magnetosphere. *Planet. Space Sci.* **49**, 283–289 (2001).
25. Smith, E. J., Davis, L. Jr & Jones, D. E. in *Jupiter* (ed. Gehrels, T.) 788–829 (Univ. Arizona Press, Tucson, 1976).
26. Leahy, D. A. *et al.* On searches for pulsed emission with application to four globular cluster x-ray sources: NGC 1851, 6441, 6624, and 6712. *Astrophys. J.* **266**, 160–170 (1983).



**Figure 3** Light-curve and power-spectrum data for the auroral hot spot. **a**, Light curve showing the X-ray count rate measured by Chandra as a function of time for the auroral hot spot. Here we defined the hot spot region to include only those X-rays emitted within a  $5^\circ$ -radius circle centred on a latitude of  $65^\circ$  and a system III longitude of  $170^\circ$  (as shown by the red circle in Fig. 2). The total number of X-rays emitted from this region is 113, and the plot shows an 11-min boxcar smoothing of a 1-min binning of the data. The orange dashed line shows the projected area of the hot spot (as a percentage of the projected area of Jupiter). The times of the HST-STIS northern auroral region images shown in Fig. 2 are indicated by vertical purple lines. Unfortunately, no images were obtained during any of the bright X-ray pulses. **b**, Power spectrum of the hot spot signal, normalized so that, if the photons were randomly distributed over the visibility period, the mean power spectral density of any particular frequency bin would be expected to have a value of 2 (ref. 26). The peak at a period of approximately 45 min is clearly seen. The peak at 300 min is associated with the approximately 600-min rotation period of Jupiter. The dashed lines are labelled with the probability of a random signal exceeding that level in a particular frequency bin (for example, the 45-min period peak has a  $4 \times 10^{-6}$  likelihood of having been attained at random).

## Acknowledgements

We thank B. H. Mauk, S. Krimigis, W. S. Kurth and M. Kaiser for helpful discussions. The support of the Chandra Project and the Smithsonian Astrophysical Observatory is gratefully acknowledged. A portion of this work is based on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc.

Correspondence and requests for materials should be addressed to G.R.G. (e-mail: randy@whistler.space.swri.edu).

# Transient aurora on Jupiter from injections of magnetospheric electrons

B. H. Mauk\*, J. T. Clarke†, D. Grodent‡, J. H. Waite Jr‡, C. P. Paranicas\* & D. J. Williams\*

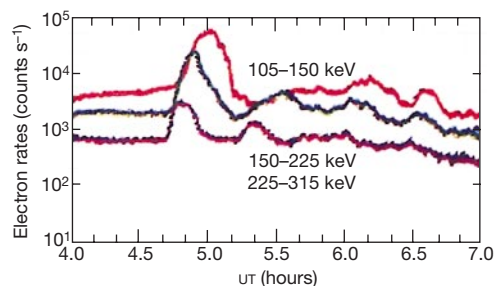
\* The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA

† Department of Astronomy and Center for Space Physics, Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

‡ Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

Energetic electrons and ions that are trapped in Earth's magnetosphere can suddenly be accelerated towards the planet<sup>1–5</sup>. Some dynamic features of Earth's aurora (the northern and southern lights) are created by the fraction of these injected particles that travels along magnetic field lines and hits the upper atmosphere<sup>4</sup>. Jupiter's aurora appears similar to Earth's in some respects; both appear as large ovals circling the poles and both show transient events<sup>6–11</sup>. But the magnetospheres of Jupiter and Earth are so different—particularly in the way they are powered—that it is not known whether the magnetospheric drivers<sup>12</sup> of Earth's aurora also cause them on Jupiter. Here we show a direct relationship between Earth-like injections of electrons in Jupiter's magnetosphere and a transient auroral feature in Jupiter's polar region. This relationship is remarkably similar to what happens at Earth, and therefore suggests that despite the large differences between planetary magnetospheres, some processes that generate aurorae are the same throughout the Solar System.

The injections within Earth's magnetosphere (Fig. 1) involve particles with kilo-electron volt (keV) to mega-electron volt



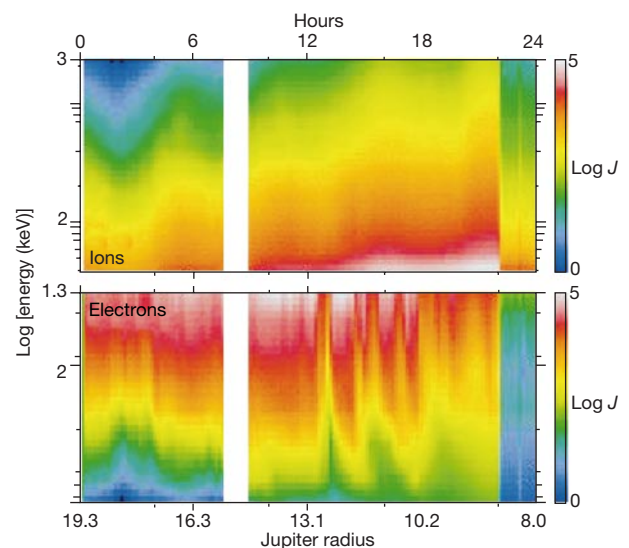
**Figure 1** Energetic electron injection measurements within Earth's space environment. The response of three different electron energy channels is shown as measured from the Earth's geosynchronous orbit (~6.7 Earth radii circular, near-equatorial). We note the energy-dispersed nature of the channels, with different energies arriving at the spacecraft at different times. Plotted after ref. 5.

(MeV) energies<sup>4</sup>. Often occurring at radial distances of 6 to 10 Earth radii, injections are one component of global dynamic events called 'magnetospheric substorms'. Substorms represent, in part, the transient release of energy stored in the magnetosphere with stressed magnetic fields<sup>13</sup>. The energy source for Earth's substorms is the solar wind of charged gases, or plasmas, emanating from the Sun. Substorms create dramatic brightening of the aurora at high geographic latitudes and a substantial expansion of the regions where auroral emissions occur.

The recent discovery of Earth-like charged particle injections within Jupiter's magnetosphere<sup>14,15</sup> is surprising because Jupiter's magnetosphere is powered mostly from the inside by the rapid but steady planetary rotation rather than from the outside by the variable solar wind. The role of injections in generating auroral emissions at Jupiter has been heretofore unknown, to our knowledge.

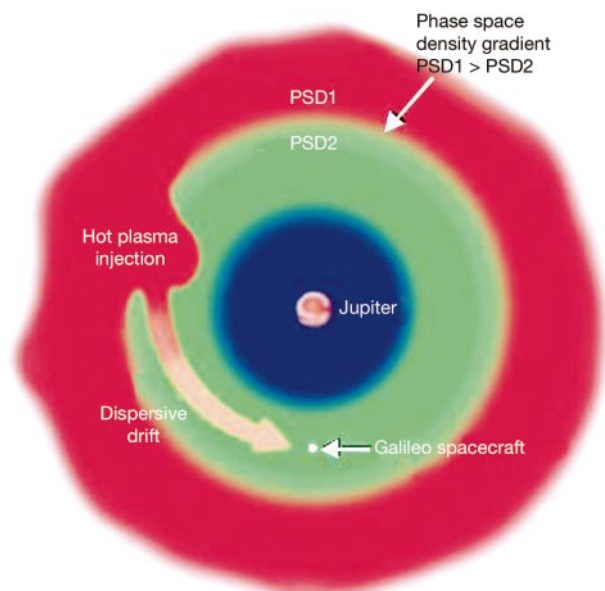
A unique opportunity to address dynamics in Jupiter's space environment was made available by a Jupiter joint observation campaign in late 2000 and early 2001. It involved the fly-by of the Cassini spacecraft, headed toward Saturn, the Galileo spacecraft orbiting Jupiter, and remote imaging by the Hubble Space Telescope (HST). During the campaign, Galileo recorded energetic electron injection signatures at radial distances of ~10 to ~13 Jupiter radii (Fig. 2). A simple model (Fig. 3) explains the energy-dispersed character of these signatures (different particle energies arrived at Galileo at different times). The model is closely analogous to models derived from injections at Earth<sup>16–18</sup>. Quantitative analysis (Fig. 4) reveals the temporal relationship between the injections and the signatures. At the times of the injections, around 15 h before the dispersed signatures were observed, Galileo was at a radial distance of about 20 Jupiter radii and recorded no obvious signature of the injections occurring closer to Jupiter.

Ultraviolet HST auroral images, similar to those in previous reports<sup>7</sup>, were taken during the Galileo operations and remapped to polar coordinates (Fig. 5). The images reveal a distinct auroral emission patch in eight consecutive images (100-s exposures)



**Figure 2** Energetic electron injection measurements within Jupiter's magnetosphere. Log [energy (keV)] versus time (hours of day 363, 2000; top scale) versus particle log [intensity ( $\text{cm}^{-1} \text{s}^{-1} \text{sr}^{-1} \text{keV}^{-1}$ ), shown as a colour scale, display of ion (top) and electron (bottom) measurements from the energetic particle detector on the Galileo spacecraft for the radial range of 19 to 8 Jupiter radii (bottom scale). The energy-dispersed injections are visible in the right-hand portion of the electron display beginning at about hour 13. The electron sensor is nearly saturated at the lower energies (top of the electron display) and so the relative variations at low energies is underrepresented here.

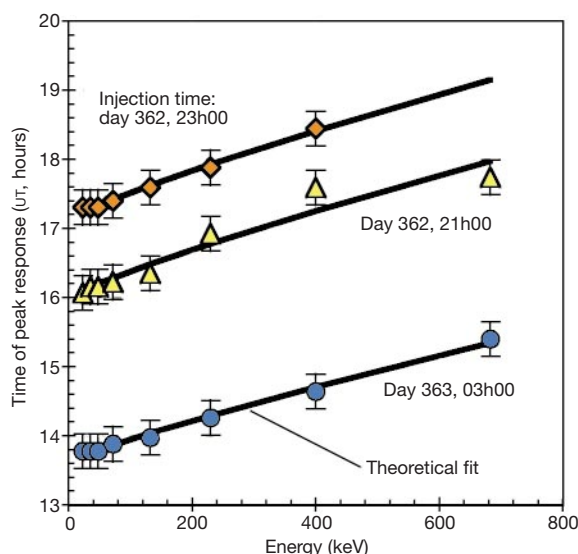




**Figure 3** Schematic for the generation of injections within Jupiter's magnetosphere. The hot plasma injection (left side) occurs quickly, and then the dispersive drift, driven by Jupiter's rotation and magnetic field inhomogeneities, occurs slowly and generates energy-dispersed particle signatures at Galileo. The phase space density (PSD) is a transformation of the particle intensities into a form that is invariant for the kind of transport thought to occur here. Plotted after refs 14 and 15.

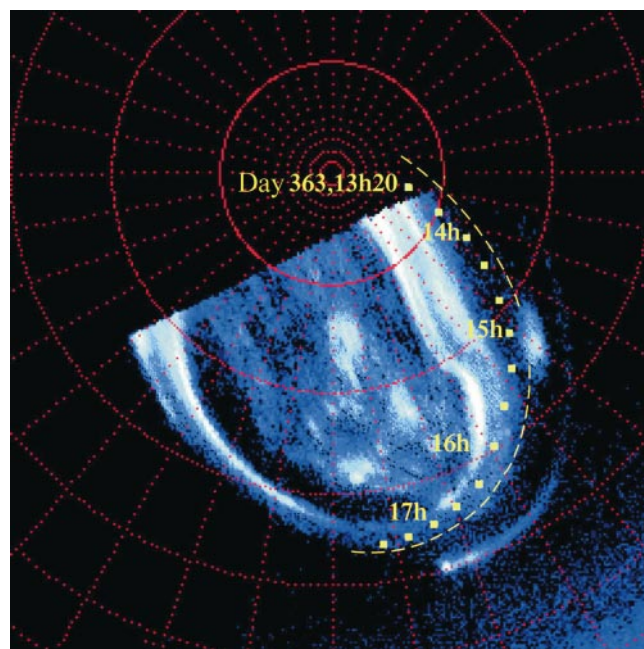
obtained over a 36-min HST operation period (Fig. 5, extreme right and close to the yellow square labelled '15 h'). The auroral patch moved with Jupiter as Jupiter rotated, and thus maintained its latitude-longitude position. The last image of the set was taken at about 1230 UT, about 30 min before the beginning of the first injection signature in Fig. 2.

We believe the patch was generated by the first of the injections



**Figure 4** Quantitative analysis of Jupiter's energy-dispersed electron injection signatures. Plotted points are the times (hours in day 363 of 2000) versus energy of the peaks in the electron channel responses for the three main electron injection signatures shown in Fig. 2. The theoretical fits use the best estimates of the energy-dependent drifts within Jupiter's magnetosphere<sup>15</sup> to reconstruct the times (shown in the figure) when the fast injections (Fig. 3) occurred. These estimates make use of a magnetic field model that incorporates electric currents internal to Jupiter and the latest estimates of typical currents external to Jupiter within the magnetosphere<sup>19</sup>.

observed by Galileo (Fig. 4). In Fig. 5, the yellow squares and the yellow dashed line show two different calculations of the motion of the Galileo spacecraft trajectory when mapped along magnetic field lines from the spacecraft to Jupiter's upper atmosphere. The yellow dashed line is the better estimate because the magnetic field model used to map Galileo's position includes contributions both from currents internal to Jupiter and from up-to-date estimates of average currents flowing external to Jupiter within its magnetosphere<sup>19</sup>. Adequate approximation to the magnetic field requires that external currents be included because the actual magnetic field configuration beyond about 9 Jupiter radii is distorted away from the nearly dipolar configuration expected from internal currents alone<sup>19,20</sup>. Projections derived using a field model with only internal currents (yellow squares in Fig. 5; VIP4<sup>21</sup>) are shown to provide a sense of the uncertainties involved with mapping. With the better estimate, the Galileo trajectory appears to have crossed into the isolated patch between ~14.7 and ~15.3 h, centred on the maximum of the higher-energy portion of the injection feature (Fig. 4). This result does not depend on detailed timing because, as mentioned, the auroral patch maintained its latitude-



**Figure 5** Hubble Space Telescope (HST) ultraviolet image of Jupiter's northern hemisphere aurora transformed to a Jupiter system-III polar coordinate system. Zero and 90 degree longitudes are straight up and horizontal to the right, respectively. The auroral emission patch of interest is to the extreme right, adjacent to the yellow square labelled '15 h'. The yellow squares and yellow dashed line show magnetic projections of the Galileo spacecraft onto Jupiter's upper atmosphere using two models of Jupiter's magnetic field (see text for details). The model of Khurana<sup>19</sup>, that includes currents external to Jupiter, was updated by replacing the internal field model, O6, with the latest internal field model, VIP4<sup>21</sup>. We note that as Galileo moved closer to Jupiter, the trajectory appears to move closer and closer to the strong emissions of the bright, poleward ring of the auroral oval. Although this is not the subject of this Letter, this characteristic is contrary to what one might expect if this ring of emission has a source that predominates steadily at some radial distance. However, as verified with previous work (see Fig. 1 of ref. 9), present field models do not predict the poleward kink in the global auroral distribution of the brightest aurora, revealed here aligned along the 140–150° longitude meridian and in previous work<sup>7</sup>. The structure of the bright auroral oval is clearly different in the kink region from its structure elsewhere, and so more than just a refinement in magnetic mapping at high latitudes will be needed to understand it. Local time effects<sup>6,9</sup> and perhaps even the effects of dynamics may be involved here. The brightest aurora is thought to map to distances as large as ~20–30 Jupiter radii, and given the mapping sensitivities<sup>9</sup>, our results are consistent with that hypothesis.

longitude position as Jupiter rotated.

The auroral patch characteristic of rotating with Jupiter is expected because the injected electrons also rotate with Jupiter (within several per cent at  $\sim 12$  Jupiter radii<sup>15</sup>). Although the auroral patch of interest here was the brightest observed during the joint observation campaign, such patches are common within HST images. Likewise, the occurrence of jovian electron injections is also common<sup>15</sup>. The other energetic electron injections revealed in the Galileo data (Figs 2 and 4) map to regions (Fig. 5) that also show measurable auroral emissions. However, those emissions do not appear patch-like and may have been active even in the absence of injections.

Studies of Earth's magnetosphere<sup>13</sup> suggest two different ways that injected energetic particles can generate auroral emissions. (1) The particle energy distributions are modified during injection and become unstable to exchanges of energy with magnetospheric wave modes. The waves scatter particles so that some travel narrowly along the magnetic field lines until they encounter the atmosphere. (2) The injected particle cloud is a high-pressure region and so electric current flows along its boundary. This pressure-driven (diamagnetic) current diverges along the leading and trailing edges of the rotating cloud because the magnetic field strength changes with radial distance. Currents are driven along the magnetic field lines towards and away from the planet and can interact strongly with plasmas close to the planet. At Jupiter, that interaction would yield downward accelerated electrons, and atmospheric auroral emissions, at the trailing edge of the rotating plasma cloud. Although there is substantial uncertainty in the magnetic mapping, the position of the auroral patch does match best with either the trailing edge of the electron cloud (the second mechanism) or the centre of that portion of the cloud that contained the higher-energy electrons measured (the first mechanism).

For an aurora resulting from the scattering process, the maximum power density that the measured ( $>20$  keV) electron cloud can provide to the aurora is  $60 \pm 30 \text{ erg cm}^{-2} \text{ s}^{-1}$ . Higher power densities are possible with the electric current generation mechanism. Models of interaction between electrons and Jupiter's atmosphere<sup>22</sup>, recalculated with the energy distribution shapes measured in Fig. 2, yield about  $3 \text{ erg cm}^{-2} \text{ s}^{-1}$  for the electron input needed to explain the auroral emissions. Thus, measured electrons can supply the requisite energy with scattering efficiencies of only 3% to 10% of the maximum. Auroral optical emission spectra were not available for this event to estimate independently the electron energies involved. However, recent Galileo auroral observations measured the tangent altitude (above the 1-bar atmospheric pressure level) of peak auroral emissions at  $245 \pm 30$  km, with some emissions extending to an altitude of  $120 \pm 40$  km (ref. 23). Atmospheric penetration of electrons modelled for diffuse aurorae require the involvement of over 48 keV electrons to explain even the peak auroral emissions<sup>24,22</sup>. □

Received 8 November 2001; accepted 1 February 2002.

- Konradi, A. Proton events in the magnetosphere associated with magnetic bays. *J. Geophys. Res.* **72**, 3829–3841 (1967).
- Arnoldy, R. L. & Chan, K. W. Particle substorms observed at the geostationary orbit. *J. Geophys. Res.* **74**, 5019–5028 (1969).
- DeForest, S. E. & McIlwain, C. E. Plasma clouds in the magnetosphere. *J. Geophys. Res.* **76**, 3587–3611 (1971).
- Mauk, B. H. & Meng, C.-I. in *Auroral Physics* (eds Meng, C.-I., Rycroft, M. J. & Frank, L. A.) 223–240 (Cambridge Univ. Press, Cambridge, 1991).
- Li, X. *et al.* Simulation of dispersionless injections and drift echos of energetic electrons associated with substorms. *Geophys. Res. Lett.* **25**, 3763–3766 (1998).
- Satoh, T., Connerney, J. E. P. & Baron, R. L. Emission source model of Jupiter's  $\text{H}_3^+$  aurorae: A generalized inverse analysis of images. *Icarus* **122**, 1–23 (1996).
- Clarke, J. T. *et al.* Far-ultraviolet imaging of Jupiter's aurora and the Io 'footprint'. *Science* **274**, 404–409 (1996).
- Ballester, G. E. *et al.* Time-resolved observations of Jupiter's far-ultraviolet aurora. *Science* **274**, 409–413 (1996).
- Satoh, T. & Connerney, J. E. P. Jupiter's  $\text{H}_3^+$  emissions viewed in corrected jovimagnetic coordinates. *Icarus* **141**, 236–252 (1999).

- Waite, J. H. *et al.* Multispectral observations of Jupiter's aurora. *Adv. Space Res.* **26**, 1453–1475 (2000).
- Waite, J. H. *et al.* An auroral flare at Jupiter. *Nature* **410**, 787–789 (2001).
- Meng, C.-I., Rycroft, M. J. & Frank, L. A. *Auroral Physics* (Cambridge Univ. Press, Cambridge, 1991).
- Kivelson, M. G. & Russell, C. T. (eds) *Introduction to Space Physics* (Cambridge Univ. Press, Cambridge, 1995).
- Mauk, B. H., Williams, D. J. & McEntire, R. W. Energy-time dispersed signatures of dynamic injections in Jupiter's inner magnetosphere. *Geophys. Res. Lett.* **24**, 2949–2953 (1997).
- Mauk, B. H. *et al.* Storm-like dynamics of Jupiter's inner magnetosphere. *J. Geophys. Res.* **104**, 22759–22778 (1999).
- McIlwain, C. E. in *Earth's Magnetospheric Processes* (ed. McCormac, B. M.) 268–279 (Reidel, Hingham, 1972).
- Konradi, A., Semar, C. L. & Fritz, T. A. Substorm-injected protons and electrons and the injection boundary model. *J. Geophys. Res.* **80**, 543–552 (1975).
- Mauk, B. H. & Meng, C.-I. Characterization of geostationary particle signatures based on the 'injection boundary model'. *J. Geophys. Res.* **88**, 3055–3071 (1983).
- Khurana, K. K. Euler potential models of Jupiter's magnetospheric field. *J. Geophys. Res.* **102**, 11295–11306 (1997).
- Connerney, J. E. P., Acuña, M. H. & Ness, N. F. Modeling the jovian current sheet and inner magnetosphere. *J. Geophys. Res.* **86**, 8370–8384 (1981).
- Connerney, J. E. P. *et al.* New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *J. Geophys. Res.* **103**, 11929–11939 (1998).
- Grodent, D., Waite, J. H. Jr & Gérard, J.-C. A self-consistent model of the jovian auroral thermal structure. *J. Geophys. Res.* **106**, 12933–12952 (2001).
- Vasavada, A. R. *et al.* Jupiter's visible aurora and Io footprint. *J. Geophys. Res.* **104**, 27133–27142 (1999).
- Ajello, J. *et al.* Galileo orbiter ultraviolet observations of Jupiter aurora. *J. Geophys. Res.* **103**, 20125–20148 (1998).

## Acknowledgements

We gratefully acknowledge help from and discussions with R. W. McEntire and T. Choo, and the support of the Space Telescope Institute.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.H.M. (e-mail: Barry.Mauk@jhuapl.edu).

# Bandgap modulation of carbon nanotubes by encapsulated metallofullerenes

Jinhwan Lee<sup>\*†</sup>, H. Kim<sup>\*†</sup>, S.-J. Kahng<sup>‡</sup>, G. Kim<sup>\*</sup>, Y.-W. Son<sup>\*</sup>, J. Ihm<sup>\*</sup>, H. Kato<sup>§</sup>, Z. W. Wang<sup>§</sup>, T. Okazaki<sup>§</sup>, H. Shinohara<sup>§</sup> & Young Kuk<sup>\*†</sup>

<sup>\*</sup> School of Physics and <sup>†</sup> Center for Science in Nanometer Scale, Seoul National University, Seoul 151-747, Korea

<sup>‡</sup> Department of Physics, Soongsil University, Seoul 156-743, Korea

<sup>§</sup> Department of Chemistry, Nagoya University, Nagoya 464-8602, Japan

Motivated by the technical and economic difficulties in further miniaturizing silicon-based transistors with the present fabrication technologies, there is a strong effort to develop alternative electronic devices, based, for example, on single molecules<sup>1,2</sup>. Recently, carbon nanotubes have been successfully used for nanometre-sized devices such as diodes<sup>3,4</sup>, transistors<sup>5,6</sup>, and random access memory cells<sup>7</sup>. Such nanotube devices are usually very long compared to silicon-based transistors. Here we report a method for dividing a semiconductor nanotube into multiple quantum dots with lengths of about 10 nm by inserting Gd@C<sub>82</sub> endohedral fullerenes. The spatial modulation of the nanotube electronic bandgap is observed with a low-temperature scanning tunnelling microscope. We find that a bandgap of  $\sim 0.5$  eV is narrowed down to  $\sim 0.1$  eV at sites where endohedral metallofullerenes are inserted. This change in bandgap can be explained by local elastic strain and charge transfer at metallofullerene sites. This technique for fabricating an array of quantum dots could be used for nano-electronics<sup>8</sup> and nano-optoelectronics<sup>9</sup>.



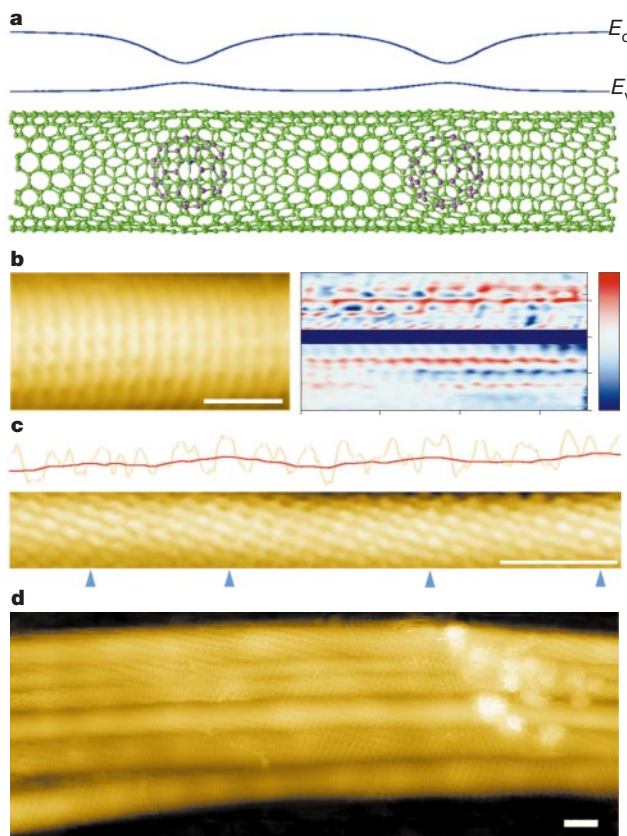
Since the discovery of carbon nanotube a decade ago<sup>10</sup>, scientists have synthesized a variety of forms of carbon nanotubes. Recently, it has been shown experimentally that fullerenes or endohedral metallofullerenes<sup>11</sup> such as Gd encapsulated inside C<sub>82</sub> (Gd@C<sub>82</sub>:GdMF) can be inserted into single-wall nanotubes (SWNTs), forming a pea-pod-like structure<sup>12–15</sup>. When the diameter of the inserted fullerene is smaller than that of the SWNT minus a certain length (~0.7 nm, roughly twice the van der Waals bond length<sup>16</sup>), the fullerenes are inserted exothermally<sup>12,18</sup>. When the diameter of a fullerene slightly exceeds the length described above, the fullerene can be inserted endothermally, but the resultant SWNT is elastically strained. A theoretical study has predicted that the electronic structure is severely modified, including the positions of the van Hove singularities (VHS), when an SWNT is uniaxially strained<sup>19</sup>. As we combine these ideas, we can perform 'local bandgap engineering' at the site where a fullerene is endothermally inserted. As schematically shown in Fig. 1a, we predict that the bandgap of a semiconducting nanotube would be altered by the insertion of large fullerenes.

The insertion of GdMFs into SWNTs was done by keeping them with SWNTs in a sealed glass ampoule at 500 °C for 3 days. The GdMFs were synthesized by a d.c. arc-discharge method<sup>11,12</sup> and isolated with a multiple-step HPLC (high-performance liquid

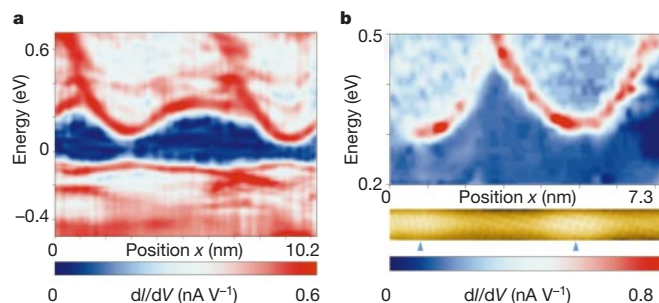
chromatography) method<sup>15</sup>. The SWNTs were generated using a d.c. arc-discharging method with a graphite-metal composite rod and treated with acids to remove amorphous carbons. Previously reported TEM images and electron energy loss spectra<sup>14,15</sup> suggested that the GdMFs can be spaced either regularly with 1.1 nm spacing or less regularly, with 1.1–3 nm spacing. The density of GdMFs could be controlled with sonication *ex situ* and/or annealing *in situ*. Figure 1b shows a typical STM image and the corresponding spatial variation of dI/dV along a SWNT without inserted fullerenes at a temperature of ~5 K. Scanning tunnelling microscopy and spectroscopy (dI/dV, where *I* is current and *V* is voltage) give bias-dependent topographic images and local density of states. To observe the local variation of the electronic structure, we took dI/dV spectra at 512 sites along the nanotube just after the topography was taken. The *x* axis indicates the position along the tube, the *y* axis the energy, and the colour scale dI/dV (shown at the right side of Fig. 1b). It takes 30–60 min to take 512 dI/dV spectra. Thermal drift is less than 0.1 nm h<sup>-1</sup> at this temperature. In the dI/dV spectra, two strong VHS peaks corresponding to the conduction and valence band edges are clearly visible with atomic resolution. As reported<sup>20,21</sup>, the conduction and valence band edges are flat. When we imaged topographies of the SWNTs with inserted GdMFs (GdMF-SWNTs) at bias voltages of -0.8 V to +1.0 V, parts of the GdMF-SWNT appeared brighter than other areas, suggesting that the diameter may be greater there than at other areas or that the local electronic structure is modified by the inserted GdMFs as shown in Fig. 1c and d. Figure 1d shows a bundle of six GdMF-SWNTs, showing the variation of spacings between neighbouring GdMFs. We note that the density is still lower than the reported TEM results<sup>12,14</sup>. About 10% of over 200 SWNT images showed locally bright spots in STM images.

Protrusions in STM topography do not necessarily indicate the exact locations of inserted GdMFs, because the electronic structures, including band edges, are severely modified with the insertion. Figure 2a shows spatial variation of dI/dV spectra along a GdMF-SWNT. The chirality of the GdMF-SWNT is determined to be (11,9) from the measured diameter of 1.4 nm (nominally 1.375 nm) and the chiral angle of 27° (nominally 26.7°). In the dI/dV spectra, conduction and valence band edges are also visible with two smaller VHS peaks following at higher bias voltages. The original bandgap of 0.43 eV is narrowed down to 0.17 eV at the sites where the GdMFs are expected to be located<sup>19</sup>. There are several causes which could account for the change in the bandgap:

(1) The magnetic field originating from the Gd atoms inside C<sub>82</sub> might cause this change. However, the magnetic flux of Gd (with the magnetic moment of ~8μ<sub>B</sub>) experienced by the SWNT is only ~10<sup>-5</sup> flux quantum, so the effect of this phase variation on the



**Figure 1** Effect of inserted Gd metallofullerenes (GdMFs) on the topography and band structure of a single-walled nanotube (SWNT). **a**, The illustration of a (11,9) GdMF-SWNT. Elastic strain is expected around the GdMFs. A schematic representation of estimated modulation of conduction ( $E_c$ ) and valence ( $E_v$ ) band edges is also shown. **b**, A typical topographic image of a SWNT obtained at the sample bias voltage of -0.3 V (left); the corresponding dI/dV spectra (right). The *x* axis represents the position along the SWNT, the *y* axis the energy in the range of -1.0 eV (bottom) to +0.87 eV (top), and the colour scale the local density of states. **c**, Atomically resolved topographic image of a (15,5) GdMF-SWNT at the sample bias voltages of 0.6 V. The GdMF sites are indicated by arrows. **d**, A bundle of six GdMF-SWNTs. Variation of spacings between neighbouring GdMFs are visible. The scale bars in all topographic images represent 1 nm.



**Figure 2** Bandgap modulation of a GdMF-SWNT by encapsulated GdMFs. **a**, The dI/dV spectra taken along the 10.2-nm-long GdMF-SWNT. The *x* axis represents the position along the GdMF-SWNT, the *y* axis the energy, and the colour scale the local density of states. **b**, Topographic image of a 7.3-nm-long (11,9) GdMF-SWNT at a sample bias voltage of 0.5 V. The sites of inserted GdMFs are indicated by arrows (bottom); the corresponding dI/dV spectra of conduction band at the centre of the GdMF-SWNT are also shown (top).



energy gap through the changed boundary condition around the SWNT is only  $10^{-3}$  times the gap ( $<0.01$  meV). Or, the energy shift of the electron on the SWNT by the magnetic field of Gd ( $\sim 100$  G) is again within 0.01 meV.

(2) The modulation of  $dI/dV$  along the GdMF-SWNT may possibly be due to the quantized electron-resonance states in a SWNT of short length or those localized between defects<sup>20,21</sup>. However, bright spots appear even in a very long GdMF-SWNT and no correlation between the observed and expected results was found. We also observed the size effect of  $dI/dV$  spectra between two defects separated by  $\sim 5$  nm in a plain SWNT<sup>22</sup>. The size effects resulted mainly in the intensity variation of VHS and the observed spatial variation of the local density of states is less than 0.1 eV.

(3) Charge transfer from the SWNT to the GdMFs or the Au(111) substrate may also induce a change in the electronic structure. The greatest effect of this mechanism is to shift the whole spectrum (both conduction and valence bands), together with some contribution to the energy-gap modification. Indeed, we notice that the Fermi level is closer to the valence band edge owing to the charge transfer, as though the SWNT were *p*-doped, and the majority of the modulation occurs in the conduction band.

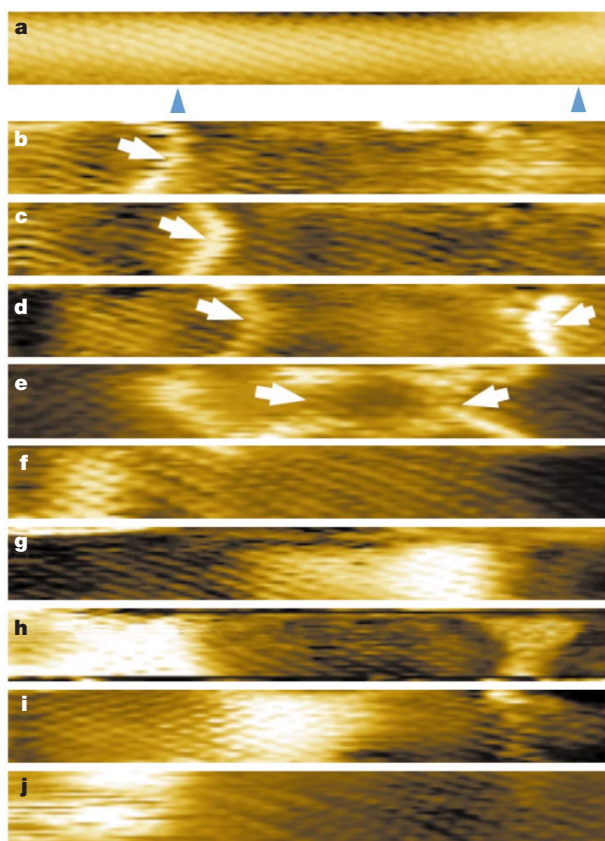
(4) As suggested above, elastic strain can change the bandgap significantly. For example, a strain of 4% in the tube axis direction can induce a gap reduction of 60% for the (15,1) GdMF-SWNT<sup>19</sup>. To study the strain produced by fullerene insertion, we performed first-principles self-consistent pseudopotential calculations using the local density approximation. The Troullier–Martins nonlocal

pseudopotential in the separable Kleinman–Bylander form is employed with an energy cutoff of 40 Rydbergs (Ry). Localized basis orbitals are used to cope with a very large unit cell. The Gd atom is either not included or replaced by a simple metal (Mg) atom because of computational limitations. The results show the same trend as in ref. 19, but the maximum calculated strain in the relaxed geometry is  $<1\%$  and the corresponding gap change is too small ( $<0.1$  eV) compared with experiment.

We therefore conjecture that the combined contribution of the elastic strain and the electron transfer to the Au substrate and to  $C_{82}$  may be responsible for the observed gap modulation. Unlike in the uniaxial strain, stretching in the circumferential direction is important here. One of the surprising features of these spectra is that the change in bandgap is so strongly localized that its spatial variation is clearly resolved in Fig. 2a. At other locations along the GdMF-SWNT, we observed the modified conduction band of a different shape as shown in Fig. 2b. Various shapes of the modified energy band reflect the variation of local elastic strain and/or charge transfer or other physical quantities not listed here. The contributions of elastic strain and charge transfer to the change in bandgaps could be studied further by repeating the same experiments on SWNTs with plain fullerenes<sup>23</sup> and other types of metallofullerenes.

With the cylindrical symmetry of the SWNT alone, we would expect that its electronic structure might also show cylindrical symmetry. However, owing to the effect of the Au substrate, the cylindrical symmetry is reduced to at most a mirror-symmetry (with respect to the normal plane that cuts the SWNT in half along the tube axis). Mirror-symmetric distributions of the elastic strain and the charge transfer may occur. For example, the closer to the substrate, the greater is the charge transfer. That would give rise to the bending of the potential and shift of the VHSs with the above-mentioned mirror symmetry. When we took  $dI/dV$  images (a spatial map of the local electronic density of states<sup>22</sup>) of a (15,5) GdMF-SWNT as a function of bias voltage, we found this expectation was correct. Figure 3a is a topographic image of a 7.6-nm-long GdMF-SWNT and Fig. 3b–j shows the corresponding  $dI/dV$  images at various bias voltages. The bandgap widening was observed at one site of this GdMF-SWNT in the tunnelling spectroscopy (not shown here). Mirror-symmetric, often sickle-shaped bright regions were observed here in Fig. 3b–j (indicated by arrows). The shift of the peaks in the local density of states marked by arrows in Fig. 3 is much greater than those of the two VHSs around the Fermi level in Fig. 2a. These peaks may be due to resonance tunnelling via unoccupied states associated with encapsulated GdMFs.

The present band structure represents one-dimensional multiple quantum dots, analogous to a multiple quantum well in a three-dimensional superlattice. The three-dimensional multi-layer utilizes the two-dimensionally confined carrier sub-bands for electronic and opto-electronic applications. The present multiple quantum dots may analogously be used for nano-optical devices. If the dots are periodic, the system can even be used for a quantum cascade laser<sup>9</sup>. With the bandgap modulation achieved here, we can create electron sub-bands in the potential-well regions. As we vary the density of the encapsulated fullerenes, the spacings between the quantum dots can be controlled. That would change the shape of the potential well for the electrons and holes. By inserting several different types of metallofullerenes, one-dimensional heterostructure, complex bandgap engineering can be achieved. With an additional electrode around the quantum wells, new one-dimensional electronic devices can be made. By using many identical quantum dots, this system could be applied to quantum computing<sup>8</sup>. We can also synthesize this bandgap-engineered system by self-assembly instead of by epitaxial growth. □



**Figure 3** Energy-resolved  $dI/dV$  images of a GdMF-SWNT. The sites of inserted GdMFs are indicated by arrows. **a**, A topographic image of a 7.6-nm-long (15,5) GdMF-SWNT. **b–j**, The corresponding  $dI/dV$  images at the biases of **b**, 1.4; **c**, 1.2; **d**, 1.0; **e**, 0.8; **f**, 0.6; **g**, 0.4; **h**, 0.2; **i**, -0.2; **j**, -0.4 V. White arrows indicate that the position of the secondary-VHS shifts along the GdMF-SWNT by changing bias voltages, showing a large dispersion. Each bright line is sickle-shaped, which indicates that charge transfer or elastic strain varies around the GdMF-SWNT, breaking the cylindrical symmetry.

Received 15 August 2001; accepted 23 January 2002.

1. Aviram, A. & Ratner, M. (eds) *Molecular Electronics: Science and Technology*. (Annals of the New York Academy of Sciences, Vol. 852, New York, 1998).

2. Reed, M. A. & Tour, J. M. Computing with molecules. *Sci. Am.* **282**(6), 86–93 (June, 2000).
3. Yao, Z., Postma, H. W. Ch., Balents, L. & Dekker, C. Carbon nanotube intramolecular junctions. *Nature* **402**, 273–276 (1999).
4. Zhou, C., Kong, J., Yenilmez, E. & Dai, H. Modulated chemical doping of individual carbon nanotubes. *Science* **290**, 1552–1555 (2000).
5. Tan, S. J., Verschuere, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
6. Postma, H. W. Ch., Tijs, T., Yao, Z., Grifoni, M. & Dekker, C. Carbon nanotube single-electron transistors at room temperature. *Science* **293**, 76–79 (2001).
7. Rueckes, T. *et al.* Carbon nanotube-based nonvolatile random access memory for molecular computing. *Science* **289**, 94–97 (2000).
8. Toth, G. & Lent, C. S. Quantum computing with quantum-dot cellular automata. *Phys. Rev. A* **63**, 052315–052323 (2001).
9. Faist, J. *et al.* Continuous wave operation of a vertical transition quantum cascade laser above  $T = 80$  K. *Appl. Phys. Lett.* **67**, 3057–3059 (1995).
10. Iijima, S. & Ichihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363**, 603 (1993).
11. Shinohara, H. Endohedral metallofullerenes. *Rep. Prog. Phys.* **63**, 843–892 (2000).
12. Suenaga, K., Iijima, S., Kato, H. & Shinohara, H. Fine structure analysis of Gd M45 near-edge EELS on the valence state of Gd@C<sub>60</sub> microcrystals. *Phys. Rev. B* **62**, 1627–1630 (2000).
13. Smith, B. W., Monthieux, M. & Luzzi, D. E. Encapsulated C<sub>60</sub> in carbon nanotubes. *Nature* **396**, 323–324 (1998).
14. Suenaga, K. *et al.* Element-selective single atom imaging. *Science* **290**, 2280–2282 (2000).
15. Hirahara, K. *et al.* One-dimensional metallofullerenes crystal generated inside single-walled carbon nanotube. *Phys. Rev. Lett.* **85**, 5384–5387 (2000).
16. Smith, B. W., Luzzi, D. E. & Achiba, Y. Tumbling atoms and evidence for charge transfer in La<sub>2</sub>@C<sub>60</sub>/SWNT. *Chem. Phys. Lett.* **331**, 137–142 (2000).
17. Okada, S., Saito, S. & Oshiyama, A. Energetics and electronic structures of encapsulated C<sub>60</sub> in a carbon nanotube. *Phys. Rev. Lett.* **86**, 3835–3838 (2001).
18. Bandow, S. *et al.* Smallest limit of tube diameters for encasing of particular fullerenes determined by radial breathing mode Raman scattering. *Chem. Phys. Lett.* **347**, 23–28 (2001).
19. Yang, L. & Han, J. Electronic structure of deformed carbon nanotubes. *Phys. Rev. Lett.* **85**, 154–157 (2000).
20. Venema, L. C. *et al.* Imaging electron wave functions of quantized energy levels in carbon nanotubes. *Science* **283**, 52–55 (1999).
21. Lemay, S. G. *et al.* Two-dimensional imaging of electronic wavefunctions in carbon nanotubes. *Nature* **412**, 617–620 (2001).
22. Kim, H., Lee, J., Kahng, S.-J. & Kuk, Y. Spatial modulation of conduction and valence band edges around defects. *Phys. Rev. Lett.* (submitted).
23. Hornbaker, D. J., Kahng, S.-J., Misra, S. & Yazdani, A. Pseudo gaps and defect states in carbon nanotubes. *Proc. 11th Int. Conf. on Scanning Tunneling Microscopy/Spectroscopy Related Techniques* 104 (National Research Council, Vancouver, Canada, 2001).

## Acknowledgements

We thank J.-Y. Park for his LTSTM design, Y. J. Song and H. S. Suh for experimental assistance, and J. J. Yu and S. W. Hwang for discussions. This work was supported by the Korean Ministry of Science and Technology through Creative Research Initiatives Program and the Future Program on New Carbon Nano-Materials by the Japan Society for the Promotion of Science.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.K. (e-mail: ykuk@phy.snu.ac.kr).

# Factors determining crystal–liquid coexistence under shear

Scott Butler & Peter Harrowell

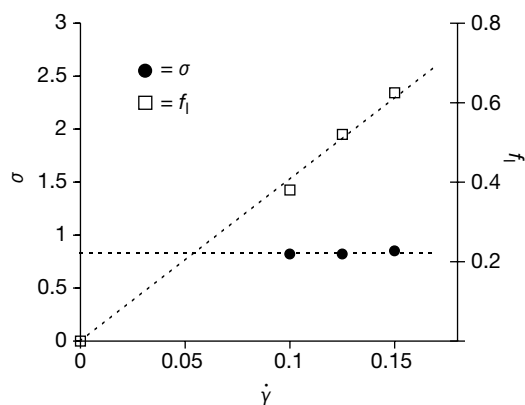
School of Chemistry, University of Sydney, Sydney, New South Wales 2006, Australia

The interaction between an imposed shear flow and an order–disorder transition underlies a broad range of phenomena. Under the influence of shear flow, a variety of soft matter<sup>1–4</sup> is observed to spontaneously form bands characterized by different local order—for example, thermotropic liquid crystals subjected to shear flow exhibit rich phase behaviour<sup>5</sup>. The stability of order under the influence of shear flow is also fundamental to understanding frictional wear<sup>6</sup> and lubrication<sup>7,8</sup>. Although there exists

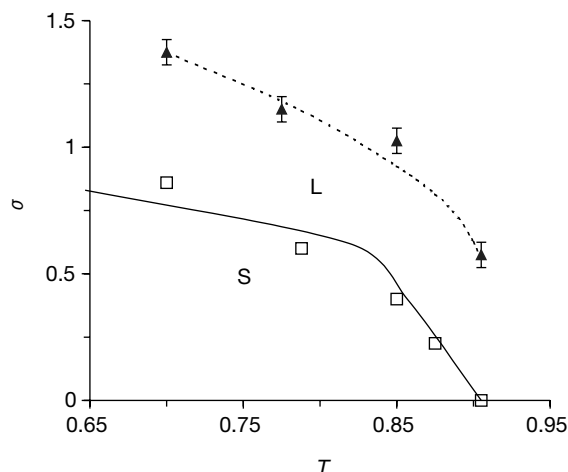
a well developed theoretical approach to the influence of shear flow on continuous transitions in fluid mixtures<sup>9</sup>, little is known about the underlying principles governing non-equilibrium coexistence between phases of different symmetry. Here we show, using non-equilibrium molecular dynamics simulations of a system of spherical particles, that a stationary coexistence exists between a strained crystal and the shearing liquid, and that this coexistence cannot be accounted for by invoking a non-equilibrium analogue of the chemical potential. Instead of such thermodynamic arguments<sup>10,11</sup>, we argue that a balancing of the crystal growth rate with the rate of surface erosion by the shearing melt can account for the observed coexistence.

Our goal is to understand the nature of the coexistence between a crystal and its melt subjected to a shear flow. Specifically, we would like to test directly the proposition that non-equilibrium coexistence involves the balancing of properties of the bulk crystal and liquid phases (just as equilibrium coexistence involves balancing the crystal and liquid chemical potentials). In order to be able to calculate such bulk properties, we shall only consider states in which both bulk phases are responding linearly to the applied shear stress. This means that we shall not consider stresses in excess of the crystal yield stress—as occurs in the case of shearing colloidal crystals<sup>12–14</sup>—nor, on the liquid side, will we consider phenomena such as shear-induced aggregation<sup>15</sup> or hydrodynamic instabilities.

We report here the results of non-equilibrium molecular dynamics simulations of 2,592 Lennard–Jones particles sheared between two parallel walls, one amorphous and one crystalline. The normal to the (111) surface of the face centred cubic (f.c.c.) crystal lies parallel to the shear gradient, with the shear flow directed along the [110] direction. We find reproducible stationary states, corresponding to coexistence between the strained crystal and its shearing melt, for a range of temperatures  $T$  below the equilibrium melting point. An example of the calculated shear stress  $\sigma$  as a function of the applied shear rate  $\dot{\gamma}$  is plotted in Fig. 1 along with the volume fraction  $f_l$  of the sample that is liquid. To determine  $f_l$ , we defined the dividing surface between crystal and liquid to be the point at which the structural order parameter  $X$  (as defined in Methods) equals 0.5. The shear stress is independent of the applied shear rate in the coexistence region, and a lever rule relates the fraction of crystal to liquid with the overall shear rate of the cell. (As required by mechanical stability, the shear stress is uniform throughout the two-phase system.) To avoid perturbations due to the walls, we have restricted our study to that part of the  $T$ – $\dot{\gamma}$  phase space for which the distance between the crystal–liquid interface



**Figure 1** Shear stress  $\sigma$  and liquid volume fraction  $f_l$  versus the applied shear rate  $\dot{\gamma}$  at  $T = 0.7$ . Note that  $\sigma$  is independent of  $\dot{\gamma}$ , whereas the liquid fraction is proportional to the applied shear rate. Although the applied shear rate  $\dot{\gamma}$  is varied over the coexistence region, the actual shear rate of the liquid remains constant.



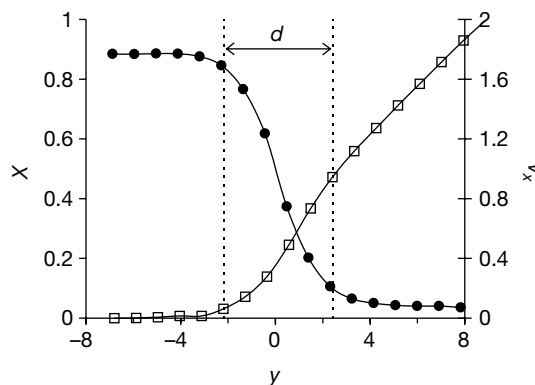
**Figure 2** Shear stress  $\sigma$  versus temperature  $T$ . Open squares, the coexistence line. Filled triangles, the yield shear stress of the bulk crystal. (The dashed line is a guide to the eye). The solid curve is the fit of the theoretical expression based on surface erosion (see equation (3)). L, liquid; S, solid.

and both walls exceeds a distance of  $\sim 5.6$  (equivalent to six (111) layers of the f.c.c. crystal). Doubling the separation between the walls and the number of particles did not result in any change in the observed coexistence states.

In the temperature–shear stress ( $T$ – $\sigma$ ) projection of the non-equilibrium phase diagram (Fig. 2), the coexistence shear stress vanishes continuously at the equilibrium melting point,  $T_M = 0.907$ , and lies well below the yield stress of the bulk crystal. (The yield stress was established by taking an equilibrated uniform crystal between two crystalline walls, and slowly increasing the applied shear stress until creep flow occurred.) We find that the suppression of the coexistence temperature is a nonlinear function of the applied stress. Simulations<sup>3,14</sup> of coexistence in shearing colloidal suspensions have indicated a linear variation of the shift in the coexistence point with respect to the applied stress; we shall return to this point later in this Letter.

Any coexistence between two phases of a single component system requires three constraints. Our simulated system satisfies equalities in temperature and stress between the two phases. At equilibrium, the extra constraint is provided by the equality of the chemical potentials of the two phases or, equivalently, the requirement that the two phase system minimizes a free energy. A number of workers<sup>10,11</sup> have suggested that this condition can be extended to non-equilibrium systems with a suitable choice of non-equilibrium chemical potential. We shall now test this suggestion.

Our choice of system proves crucial to addressing this question. The crystal strain exhibits a linear relation with stress at all the coexistence points studied. It follows that the strain-induced increase in the crystal chemical potential  $\mu_S$  can be accurately



**Figure 3** The crystal structure  $X$  (filled circles) and shear velocity  $v_x$  (open squares). These quantities have been averaged over layers parallel to the interface along  $y$ , the shear gradient direction at  $T = 0.7$ . The origin along the surface normal was defined as the point at which the crystal structure profile  $X = 0.5$ . The average shear velocity  $v_x$  is measured with respect to the crystalline wall.

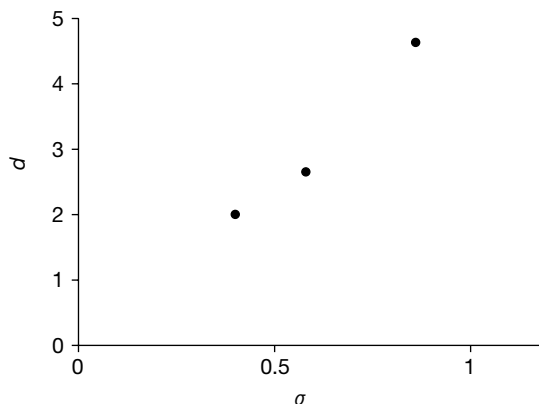
calculated using the equation

$$\Delta\mu_{\text{elastic}} = \mu_S(T, \sigma) - \mu_S(T, 0) = \frac{\sigma^2}{2\rho c_4(T)} \quad (1)$$

where  $\rho$  is the crystal density and  $c_4$  is the shear elastic coefficient ( $c_4 = d\sigma/d\gamma$  for small values of the strain  $\gamma$  in the homogeneous crystal). Over the temperature range of interest, we find  $\rho(T) \approx 1.0319 - 0.06215T$  and  $c_4(T) \approx 35.101 - 22.062T$ . Based on the absence of any significant dependence of the crystal–liquid coexistence on the nature of the thermostat (we have studied wall thermostats as well as the homogeneous thermostat described here), we argue that heat flux contributions to the chemical potential are negligible at these low shear rates. We can also accurately calculate the chemical potential difference between the unstrained crystal and the unstrained liquid using the expansion

$$\Delta\mu(T) = \mu_S(T, 0) - \mu_L(T, 0) = \frac{\Delta h(T_M)}{T_M} \Delta T + O(\Delta T^2) \quad (2)$$

where  $\Delta h$  is the heat of fusion per particle at the melting temperature  $T_M$  and  $\mu_L(T, 0)$  is the chemical potential of the unstrained liquid. We find that  $\Delta h = -1.37$  for  $T_M = 0.907$  in our system. (In calculations of the enthalpy of the bulk crystal and liquid, we used two crystalline walls or two amorphous walls, respectively.) By adding  $\Delta\mu_{\text{elastic}}$  to  $\Delta\mu(T)$  we can calculate the chemical potential difference between the strained crystal and the unstrained supercooled liquid for the temperatures studied here (see Table 1). We



**Figure 4** The penetration depth  $d$  of the shear flow into the crystal interface (as defined in the text) versus the shear stress at coexistence.

**Table 1** Chemical potential differences

| $T$   | $\Delta\mu_{\text{elastic}}$ | $\Delta\mu(T)$ | $\Delta\mu(T) + \Delta\mu_{\text{elastic}}$ |
|-------|------------------------------|----------------|---------------------------------------------|
| 0.7   | 0.019                        | −0.313         | −0.294                                      |
| 0.788 | 0.010                        | −0.180         | −0.170                                      |
| 0.85  | 0.005                        | −0.087         | −0.082                                      |
| 0.875 | 0.002                        | −0.049         | −0.047                                      |
| 0.907 | 0.0                          | 0.0            | 0.0                                         |

Chemical potential differences  $\Delta\mu(T) + \Delta\mu_{\text{elastic}}$  between the strained crystal and the unstrained supercooled liquid for  $T \leq 0.907$ , the equilibrium melting point. The increase in chemical potential for the crystal due to elastic strain  $\Delta\mu_{\text{elastic}}$  as calculated using equation (1). The chemical potential difference between the unstrained solid and unstrained liquid  $\Delta\mu(T)$  was calculated using equation (2).



find that the chemical potential of the strained crystal lies significantly below that of the unstrained supercooled liquid. As any non-equilibrium contribution to the liquid chemical potential would be expected to further increase the liquid potential (that is, moving it away from that of the solid), we conclude that there exists no physically reasonable definition of a non-equilibrium liquid chemical potential that would equal the chemical potential of the crystal. This is a significant result, as it excludes a variational condition for the stationary coexistence based on a non-equilibrium analogue of a free energy.

If the extension of thermodynamics based on equating properties of separate phases fails, we are left to conclude that the coexistence is established by the interface. We propose that the coexistence results from the balancing of two opposing kinetic processes at the crystal–liquid interface. On one hand, the crystal is growing into the supercooled liquid, driven by the chemical potential difference. As noted above, the straining of the crystal does little to reduce this difference. On the other hand, the action of the shearing liquid is to mechanically disrupt the fragile liquid side of the crystal–liquid interface. We shall refer to this process as erosion. (Ajdari<sup>16</sup> has presented a phenomenological model of the related process involving shear disruption of aggregates at a wall.) We note that the erosion of a solid by the flow of its melt described here occurs at low Reynolds numbers, characterized by simple lamellar flow in the liquid.

We are unaware of a molecular-level description of erosion, and so we shall propose here a simple treatment. We shall assume that the local rate at which crystalline order is disrupted by shear flow is proportional to the product of the local degree of crystalline order and the local strain rate: that is,  $X\dot{\gamma}_{\text{local}}$ . Erosion, therefore, involves the penetration of the shear flow into the crystal. This penetration is clearly seen in Fig. 3 where we plot the average profiles of the crystal order parameter  $X$  and the local shear velocity  $v_x$  through the stationary interface. We propose that the overall rate of erosion of the interface is simply the average rate of crystal disruption through the interface, and so is proportional to  $\int_{-\infty}^{\infty} dy X(y) \dot{\gamma}_{\text{local}}(y)$ .

The rate of crystal growth at low supercoolings is proportional to the supercooling  $\Delta T$ . In the spirit of the simple separation of the growth and erosion processes proposed above, we shall assume that this crystallization rate can be regarded as independent of the applied shear stress. Equating the rate of erosion with the rate of crystallization results in the following relation for the non-equilibrium coexistence temperature:

$$T = T_M - A \int_{-\infty}^{\infty} dy X(y) \dot{\gamma}_{\text{local}}(y) \quad (3)$$

where  $A$  is an undetermined positive constant. Adjusting  $A$  to optimize the fit, we have plotted the curve obtained from equation (3) in Fig. 2 and find that it provides a reasonable representation of dependence of the non-equilibrium coexistence temperature on the applied shear stress. This result supports our proposal that it is the coupling of the fragile interfacial structure to the imposed shear that dominates this non-equilibrium coexistence. Olmsted and Lu<sup>17</sup> recently reached a similar conclusion regarding the importance of heterogeneities in determining non-equilibrium coexistence in the case of rigid rod suspensions. Having argued that coexistence is determined by the coupling of the shear flow with the interfacial structure, it follows that details affecting this coupling—such as the direction of shear flow with respect to the crystal lattice and the spatial extent of the interface—may significantly influence the observed coexistence.

To understand how the non-equilibrium crystal–liquid interface varies along the coexistence line, we define the penetration depth  $d$  to be the distance between the points along the surface normal at which the crystalline order parameter equals 0.1 and at which the shear velocity relative to the crystalline wall drops to 0.06 (see Fig. 3). A plot of the penetration depth  $d$  versus the shear stress is shown in

Fig. 4. We find that the interface of the Lennard–Jones crystal is penetrated to a depth that increases substantially as the shear stress is increased. These results, along with examination of the structural and velocity profiles through the interface, indicate that the non-equilibrium crystal–liquid interface includes a band of shearing crystal whose thickness increases rapidly with shear stress. The previous simulations of the non-equilibrium phase diagram of a colloidal crystal<sup>13,14</sup> considered the coexistence between a shearing crystal and its disordered phase. As the penetration depth  $d$  increases with the applied shear stress, we expect our simulations eventually to resemble these earlier results for the colloidal system. □

## Methods

The walls are selected to align the crystal–liquid interface with the imposed shear flow. One wall consists of three (111) layers of an f.c.c. crystal of Lennard–Jones particles. The wall structure is chosen to be identical to the crystal structure of the particles. The other wall consists of an equivalent number (that is, 432) of the same particles in an amorphous configuration. We have found that this rough amorphous wall leaves the density of the adjacent liquid unperturbed—we see no layering of the liquid normal to this wall. The particle–particle and particle–wall interactions are chosen to be the same Lennard–Jones potential  $\phi(r) = 4\epsilon[(a/r)^{12} - (a/r)^6]$  truncated at a distance of  $2.7a$ . The interaction between the wall and the particles is the same as that between the particles. The particles in both walls are maintained in their positions by harmonic tethers with a force constant of  $57.1 \epsilon/a^2$ . A constant shear rate is applied by moving each wall a displacement  $\Delta x = 0.5\dot{\gamma}L_y\Delta t$  per time step. Note that the flow is along the  $x$  axis and the shear gradient is along the  $y$  axis. Here  $\dot{\gamma}$  is the applied shear rate,  $L_y$  is the distance between the walls, and the time step  $\Delta t = 0.004$ . (All quantities are reported here in reduced units, such that the particle mass, the length  $a$  and the energy  $\epsilon$  all equal one.)

The normal pressure is maintained at a fixed value of 3.5 by adjusting the wall separation  $L_y$  via a Nose–Hoover constraint<sup>18</sup>. The temperature is constrained by dividing the liquid into layers parallel to the walls and, for each layer, rescaling the components of the particle momenta along the shear gradient and the vorticity<sup>19</sup>. We have also studied the effect of temperature gradient arising when wall thermostats only are used. These results will be reported elsewhere (S.B. and P.H., manuscript in preparation). Crystal structure is measured along the normal to the crystal–liquid interface by calculating the in-plane hexagonal order  $X(l)$  in the  $l$ th layer in the  $xz$  plane (with a layer thickness equal to the spacing between (111) planes in the f.c.c. crystal), given by

$$X(l) = \frac{1}{N_l} \sum_{i=1}^{N_l} \frac{1}{N_i(N_i-1)} \sum_{j=1}^{N_i} \sum_{k \neq j}^{N_i} \cos(6\theta_{ijk}) \quad (4)$$

where  $\theta_{ijk}$  is the angle in the  $xz$  plane between the vectors  $\mathbf{r}_{ij}$  and  $\mathbf{r}_{ik}$ , the vectors joining the centres of particle  $i$  and two of its nearest neighbours, particles  $j$  and  $k$ .  $N_i$  is the number of nearest neighbours of particle  $i$  and  $N_l$  is the total number of particles in the  $l$ th layer.

Received 15 May; accepted 27 December 2001.

1. Berret, J., Roux, D. & Porte, G. Inhomogeneous shear flows of wormlike micelles: A master dynamic phase diagram. *Phys. Rev. E* **55**, 1668–1676 (1997).
2. Eiser, E., Molino, F., Porte, G. & Diat, O. Nonhomogeneous textures and banded flow in a soft cubic phase under shear. *Phys. Rev. E* **61**, 6759–6764 (2000).
3. Chen, L., Chow, M., Ackerson, B. & Zukoski, C. Rheological and microstructural transitions in colloidal crystals. *Langmuir* **10**, 2817–2829 (1994).
4. Fischer, E. & Callaghan, P. T. Shear banding and the isotropic–nematic transition in wormlike micelles. *Phys. Rev. E* **64**, 011501–011516 (2001).
5. Tsuji, T. & Rey, A. The effect of long range order on sheared liquid crystalline materials. *Phys. Rev. E* **62**, 8141–8151 (2000).
6. Sasada, T. & Mishina, H. Freezing point depression within a shear field, with special reference to metallurgical transformations occurring during wear. *Wear* **74**, 263–274 (1981).
7. Thompson, P. A. & Robbins, M. O. Origin of stick-slip motion in boundary lubrication. *Science* **250**, 792–794 (1990).
8. Bordarier, P., Schoen, M. & Fuchs, A. Stick-slip phase transitions in confined solid-like films from an equilibrium perspective. *Phys. Rev. E* **57**, 1621–1635 (1998).
9. Onuki, A. Phase transitions of fluids in shear flow. *J. Phys. Condens. Matter* **9**, 6119–6157 (1997).
10. Ramaswamy, S. & Renn, S. Theory of shear-induced melting of colloidal crystals. *Phys. Rev. Lett.* **56**, 945–948 (1986).
11. Baranyai, A. & Cummings, P. Towards a computational chemical potential for nonequilibrium steady-state systems. *Phys. Rev. E* **60**, 5522–5527 (1999).
12. Ackerson, B. J. & Clarke, N. A. Shear-induced partial translational ordering of a colloidal solid. *Phys. Rev. A* **30**, 906–918 (1984).
13. Stevens, M. & Robbins, M. Simulations of shear-induced melting and ordering. *Phys. Rev. E* **48**, 3778–3792 (1993).
14. Butler, S. & Harrowell, P. The shear induced disordering transition in colloidal crystals: Non-equilibrium Brownian dynamics simulations. *J. Chem. Phys.* **103**, 4653–4671 (1995).
15. Goveas, J. & Pines, D. J. A phenomenological model of shear-thickening in wormlike micelle solutions. *Europhys. Lett.* **48**, 706–712 (1999).
16. Ajdari, A. Rheological behaviour of a solution of particles aggregating on the containing walls. *Phys. Rev. E* **58**, 6294–6298 (1998).
17. Olmsted, P. & Lu, C.-Y. D. Phase separation of rigid-rod suspensions in shear flow. *Phys. Rev. E* **60**, 4397–4415 (1999).

18. Allen, M. & Tildesley, D. *Computer Simulations of Liquids* (Clarendon, Oxford, 1990).  
 19. Stevens, M. *et al.* Comparison of shear flow hexadecane in a confined geometry and in bulk. *J. Chem. Phys.* **106**, 7303–7314 (1997).

# Acknowledgements

This work was supported by an Institutional Grant from the Australian Research Council.

Correspondence and requests for materials should be addressed to P.H.  
 (e-mail: peter@chem.usyd.edu.au).

# High mixing rates in the abyssal Southern Ocean

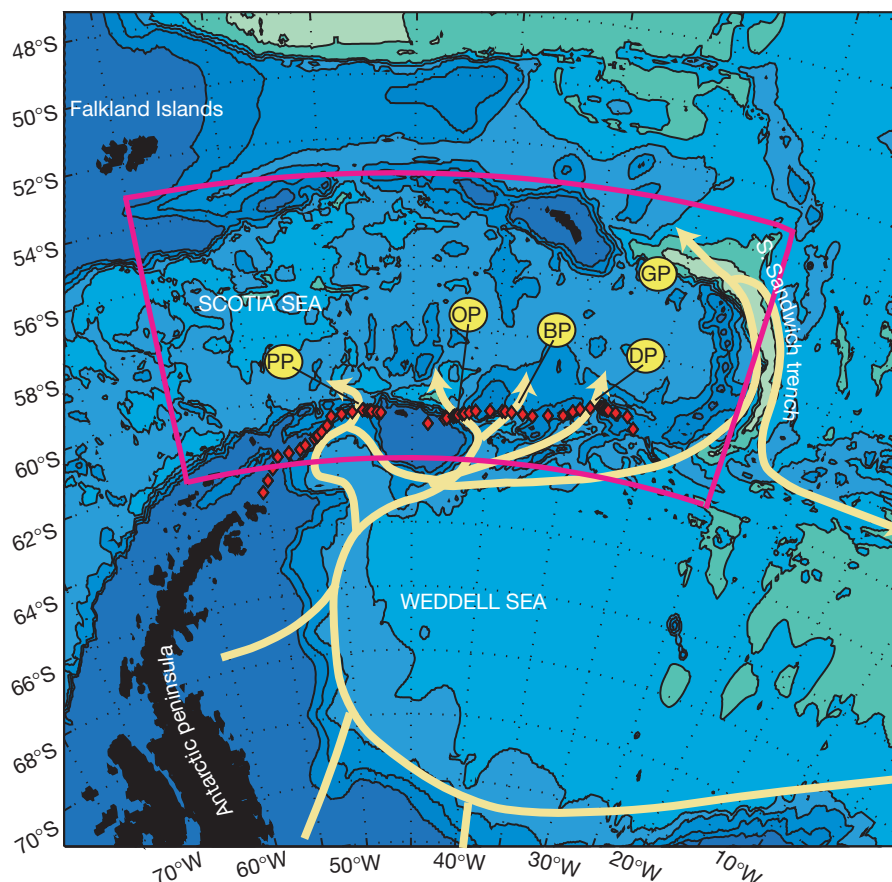
Karen J. Heywood\*, Alberto C. Naveira Garabato\* & David P. Stevens†

\* School of Environmental Sciences, † School of Mathematics,  
 University of East Anglia, Norwich NR4 7TJ, UK

Mixing of water masses from the deep ocean to the layers above can be estimated from considerations of continuity in the global ocean overturning circulation<sup>1–3</sup>. But averaged over ocean basins, diffusivity has been observed to be too small<sup>4–12</sup> to account for the global upward flux of water, and high mixing intensities have only been found in the restricted areas close to sills and narrow

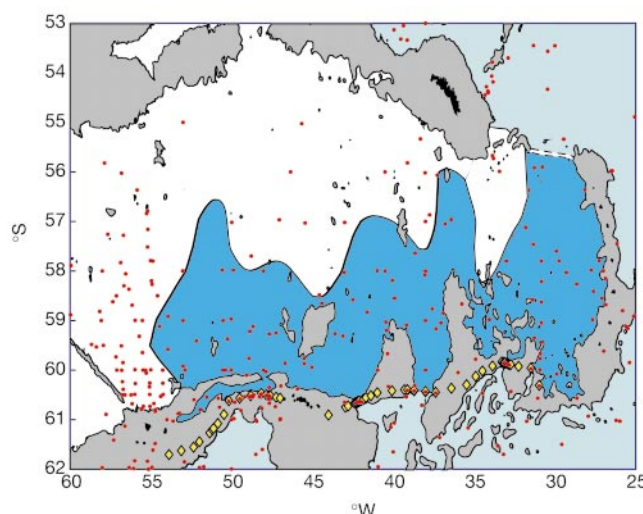
gaps<sup>10,11,13–15</sup>. Here we present observations from the Scotia Sea, a deep ocean basin between the Antarctic peninsula and the tip of South America, showing a high intensity of mixing that is unprecedented over such a large area. Using a budget calculation over the whole basin, we find a diffusivity of  $(39 \pm 10) \times 10^4 \text{ m}^2 \text{ s}^{-1}$ , averaged over an area of  $7 \times 10^5 \text{ km}^2$ . The Scotia Sea is a basin with a rough topography<sup>16</sup>, situated just east of the Drake passage where the strong flow of the Antarctic Circumpolar Current is constricted in width. The high basin-wide mixing intensity in this area of the Southern Ocean may help resolve the question of where the abyssal water masses are mixed towards the surface.

Diapycnal eddy diffusivity  $K$  is a fundamental property of ocean flows, where turbulence and mixing dominate over molecular-scale diffusion. It quantifies the exchange of heat, salt and other tracers. Model evidence suggests that the value of  $K$  determines the strength of the overturning circulation<sup>17,18</sup> that controls the climate over much of the Earth. Yet the value of  $K$  is still remarkably unbounded over orders of magnitude<sup>3</sup>. An order of magnitude of  $10^{-4} \text{ m}^2 \text{ s}^{-1}$  for the globally averaged value of  $K$  has been obtained by estimating that approximately 25 Sv ( $1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$ ) of cold, dense abyssal water forms in polar regions, and by assuming for simplicity that it upwells uniformly elsewhere into the thermocline<sup>1,2</sup>. More sophisticated versions of this calculation using inverse techniques and continent-to-continent hydrographic sections reveal basin-averaged diffusivities of similar magnitude, ranging from  $(3–4) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  for mid-depths to  $(9–12) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  for bottom layers<sup>19</sup>. Yet direct measurements suggest that  $K$  in most regions of the world



**Figure 1** Map of the Scotia Sea, showing ALBATROSS stations (red diamonds) along the South Scotia Ridge. Bathymetric contours are shown every 1,000 m. Land is shaded black. Schematic pathways of Weddell Sea Deep Water (WSDW) are shown (yellow lines). Important passages through bathymetric constraints are marked: Georgia passage (GP) through the northern boundary of the Scotia Sea, and Phillip passage (PP), Orkney passage (OP), Bruce passage (BP) and Discovery passage (DP) through the South Scotia Ridge. The pink box marks the area shown in Fig. 2.

through the northern boundary of the Scotia Sea, and Phillip passage (PP), Orkney passage (OP), Bruce passage (BP) and Discovery passage (DP) through the South Scotia Ridge. The pink box marks the area shown in Fig. 2.



**Figure 2** Map of the extent of the  $\gamma^n = 28.31 \text{ kg m}^{-3}$  neutral density surface, shaded blue and light green. The blue area shows the closed area considered for the calculation of diffusivity. Red dots denote locations of all available historical hydrographic stations that

contain WSDW. ALBATROSS stations (yellow diamonds) are shown along the South Scotia Ridge. Depths shallower than 2,000 m are shaded grey; land is shaded black. The dashed lines in the Georgia passage indicate the absence of LWSW at the sill.

ocean is an order of magnitude less than this<sup>4</sup>. Recently it has been argued<sup>20</sup> that wind-driven upwelling in the Southern Ocean reduces the requirement for global average diapycnal eddy diffusivity to  $3 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$ . However, this still demands some regions of substantially higher diffusivity.

The ALBATROSS (Antarctic large-scale box analysis and the role of the Scotia Sea) project<sup>21</sup> was designed to quantify the flow of water into and out of the Scotia Sea, particularly the deep, dense water that is formed in the Weddell Sea and provides the cold layers at the bottom of much of the world ocean. A quasi-synoptic box of densely spaced stations measuring temperature, salinity and current velocity was completed around the Scotia Sea (Fig. 1), mostly along the crest of the ridges surrounding the abyssal basin. Here we focus on the stations along the top of the South Scotia Ridge. Weddell Sea Deep Water (WSDW), with potential temperatures between  $-0.7$  and  $0^\circ\text{C}$  and neutral density  $\gamma^n$  (ref. 22) between  $28.40$  and  $28.26 \text{ kg m}^{-3}$ , forms in the Weddell Sea and escapes to the north, either through gaps in the South Scotia Ridge, or along the South Sandwich trench and the deep abyssal plains to the east<sup>23</sup>. It is the densest water escaping the Weddell Sea, and forms part of what is known as Antarctic Bottom Water, as it pervades the world ocean.

From ALBATROSS, concurrent measurements of velocity and water properties are available across each of the deep sills through which deep water can escape the Weddell Sea into the Scotia Sea: the Orkney passage, Philip passage, Bruce passage and Discovery passage (Fig. 2). The densest waters in the bottom of the Scotia

Sea emanate from the Weddell Sea through these gaps in the South Scotia Ridge<sup>23</sup>. The class of WSDW colder than  $-0.25^\circ\text{C}$  and denser than  $\gamma^n = 28.31 \text{ kg m}^{-3}$  (known as Lower WSDW, LWSW<sup>24</sup>) cannot escape the Scotia Sea, as it is too deep to flow through the Georgia passage and there are no other inflows or outflows for waters of these densities. Thus our starting point is that the water denser than  $\gamma^n = 28.31 \text{ kg m}^{-3}$  has entered the basin through the deep gaps in the South Scotia Ridge, and can only escape through mixing with overlying waters. By undertaking a heat budget of this water, we can deduce  $K$ .

If the temperature in the deep ocean is in steady state, the upward advection of cold water must be balanced by a downward diffusive heat flux<sup>1</sup>. For a flow into an enclosed basin, the heat budget can be written as

$$\rho V C_p (\theta_u - \theta_i) = A \left( G - \rho C_p K \frac{\partial \theta}{\partial z} \right)$$

where  $\rho$  is the *in situ* density of sea water;  $V$  is the volume transport of water into the basin;  $C_p$  is the specific heat capacity of sea water;  $\theta_i$  and  $\theta_u$  are the potential temperatures of the inflowing water and of that upwelled through the upper surface;  $A$  is the area of the basin;  $G$  is the geothermal heat flux;  $K$  is the thermal diapycnal eddy diffusivity; and  $\partial \theta / \partial z$  is the mean vertical gradient of potential temperature.

Values used for the calculation of  $K$  are summarized in Table 1, and their derivation discussed in the Methods section. Diapycnal

**Table 1** Parameters used in the calculation of diapycnal eddy diffusivity  $K$

|                                                                                                                                       | Heat budget                                                      | Salt budget                                             |
|---------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------|
| $\rho$ , <i>in situ</i> density of sea water                                                                                          | $1,040 \text{ kg m}^{-3}$                                        | Not required                                            |
| $V$ , volume transport of water into basin                                                                                            | $4.0 \pm 0.5 \text{ Sv}^*$                                       | $4.0 \pm 0.5 \text{ Sv}$                                |
| $\theta_i$ , or $S_i$ , potential temperature or salinity of inflowing water                                                          | $-0.438 \pm 0.007^\circ\text{C}$                                 | $34.654 \pm 0.002$                                      |
| $\theta_u$ , or $S_u$ , potential temperature or salinity of water upwelled through upper surface                                     | $-0.250 \pm 0.003^\circ\text{C}$                                 | $34.663 \pm 0.003$                                      |
| $C_p$ , specific heat capacity of sea water                                                                                           | $3.91 \times 10^3 \text{ J kg}^{-1} \text{ }^\circ\text{C}^{-1}$ | Not required                                            |
| $A$ , area of basin                                                                                                                   | $(7.0 \pm 1.0) \times 10^{11} \text{ m}^2$                       | $(7.0 \pm 1.0) \times 10^{11} \text{ m}^2$              |
| $G$ , geothermal heat flux                                                                                                            | $0.1 \pm 0.1 \text{ W m}^{-2}$                                   | Not required                                            |
| $\frac{\partial \theta}{\partial z}$ or $\frac{\partial S}{\partial z}$ , mean vertical gradient of potential temperature or salinity | $(2.7 \pm 0.3) \times 10^{-4} \text{ }^\circ\text{C m}^{-1}$     | $(6.8 \pm 0.5) \times 10^{-6} \text{ m}^{-1}$           |
| $K$ , diapycnal eddy diffusivity                                                                                                      | $(39 \pm 10) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$          | $(79 \pm 45) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ |

See Methods for details of calculations.

\*  $1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$ .



eddy diffusivity in the Scotia Sea is calculated to be  $39 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ . Accounting for each of the error sources (Table 1), and assuming that they are independent, gives an uncertainty of  $10 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ . A similar calculation can be made using a salinity budget (Table 1), yielding a value for  $K$  of  $(79 \pm 45) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ . We assume that there is no loss or gain of salt at the sea bed. The uncertainty in  $K$  is much larger because the difference between inflowing and upwelling salinity is small compared with the achievable precision of salinity measurements. That the two estimates of  $K$  are indistinguishable, within the uncertainties, indicates that the dominant process here is diapycnal mixing by fine-scale turbulence<sup>9</sup>.

The vertical length scale  $L$  ( $L = KA/V$ ) over which the upwelling water mixes is  $(680 \pm 180) \text{ m}$ . As recently pointed out<sup>20</sup>, to close the global overturning circulation, the Antarctic Bottom Water need only be mixed to the depth of the deep waters sourced in the North Atlantic, and not to the sea surface. In the Scotia Sea, our data show that this mixing is indeed accomplished. The volume of LWDSW in the Scotia Sea is  $(3.0 \pm 0.5) \times 10^{14} \text{ m}^3$ , which implies a residence time of only  $(2.4 \pm 0.5) \text{ years}$ , a remarkably rapid flushing time.

Previous budget studies using similar calculations for other abyssal basins have not found such a high value as we have for the Scotia Sea. These include  $4 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ ,  $(3.6\text{--}7.2) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  and  $(3 \pm 2) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  averaged over the Brazil basin<sup>6,9,12</sup>,  $(1\text{--}4) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  for North Atlantic abyssal basins<sup>3,7</sup> and  $(10.6 \pm 2.7) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  for the Somali basin in the Indian Ocean<sup>8</sup>. The Somali and Brazil basins are comparable in area to the Scotia Sea basin considered here. More recently, attention has been focused on regions in the close vicinity of sills and gaps through ridges. Direct measurements<sup>13</sup> of turbulent mixing of Antarctic Bottom Water in and just downstream of the Romanche fracture zone give  $K$  as high as  $150 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ , suggesting that significant mixing occurs within narrow passages. Using similar methods to ours, in and just downstream of the Samoa passage diffusivities were determined of the order of  $(50\text{--}500) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  depending on the values taken for the vertical gradient of potential temperature<sup>10</sup>. Diffusivity in the southern central Indian basin averages  $(35 \pm 14) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ , ranging from  $105 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  close to the sill in the Ninetyeast Ridge, and  $13 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$  away from the sill<sup>11</sup>. The area of our diffusivity calculation is more than an order of magnitude greater than the areas considered by each of these authors  $((2\text{--}3) \times 10^4 \text{ km}^2)$  compared with our area of  $7 \times 10^5 \text{ km}^2$ . We note that the high value of  $K$  calculated here for the Scotia Sea is a basin average. The Scotia Sea is a region of rough bathymetry, but the value was not determined for a constricted passage.

The LWSDW exported to the Scotia Sea represents up to 38% of the total WSDW exported from the Weddell Sea<sup>23</sup>. If the Weddell Sea contributes 60–70% of all Antarctic Bottom Water production<sup>25</sup>, defined as the global ocean waters denser than  $\gamma^n = 28.27 \text{ kg m}^{-3}$ , then the LWSDW upwelling in the Scotia Sea represents up to 20% of all Antarctic Bottom Water production. This is a significant fraction of the water mass that cools the global ocean, and implies that ocean models that miss the exchange of waters between the Weddell and Scotia Sea—by not resolving the flow through the deep, narrow passages in the ridges around the Scotia Sea—will be unable to represent accurately the lower limb of the overturning circulation.

The Southern Ocean is generally weakly stratified, shown by the low value of Brunt-Väisälä frequency of  $(5.4 \pm 0.2) \times 10^{-4} \text{ s}^{-1}$  at the  $\gamma^n = 28.31 \text{ kg m}^{-3}$  isopycnal in the Scotia Sea. A high value of diffusivity does not necessarily imply a large energy dissipation, if stratification is low. The dissipation we find is  $1.1 \times 10^{-9} \text{ W kg}^{-1}$ , similar to, but slightly smaller than, values deduced for the deep Brazil basin<sup>12</sup>. Taking the median thickness of the LWSDW layer to be  $(423 \pm 20) \text{ m}$ , to mix the LWSDW into the overlying waters implies an energy dissipation of  $(1.8 \pm 0.5) \times 10^9 \text{ W}$ : small in

comparison with the global requirements of  $2.1 \times 10^{12} \text{ W}$  (ref. 2) or  $0.6 \times 10^{12} \text{ W}$  (ref. 20), and not disproportionate with the area over which the mixing is occurring.

Direct measurements of abyssal turbulence<sup>14</sup> indicate that mixing rates in the bulk of the ocean interior are at least an order of magnitude smaller than the basin-averaged budget-derived values. Enhanced diffusivity has been observed over steeply sloping boundaries and seamounts in the abyssal ocean:  $K > 1 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ , compared with values an order of magnitude smaller over smooth topography<sup>4,14,15</sup>. Our very large value for  $K$  suggests that the rough topography of the Scotia Sea plays a role in enhancing the mixing in this region. It is possible that the mechanism for this is the interaction of internal tides or other internal waves with topography<sup>2,15</sup>, such as internal wave reflection occurring at sloping boundaries. Global maps of the energy estimated to be dissipated by tides<sup>26,27</sup> imply that the Scotia Sea may be a region where large amounts of energy are supplied to internal tides by the tidal flow of a stratified ocean over rough topography. Other possible sources for the energy are internal lee waves generated by interaction of the currents with the rough topography<sup>28</sup>, and basin-trapped barotropic planetary wave modes<sup>29</sup>.

The work that we report here leads us to believe that the Scotia Sea acts as a 'blender' to mix ocean water masses, and that *in situ* observations of turbulence in this region would confirm a significant contribution to the 'missing mixing' required to balance the global ocean overturning circulation. □

## Methods

### Inflow volume transport and inflow properties

LADCP (lowered acoustic Doppler current profiler) measurements throughout the water column provide a known level with which to reference geostrophic shear<sup>23</sup>. Bottom triangles below the deepest common depth of adjacent stations are included by extrapolating the deepest velocity value to the sea bed. An estimate of the uncertainty in the transport is provided from 1,000 calculations each applying a random barotropic perturbation to each LADCP profile, assuming a normal distribution with a standard deviation of  $3 \text{ cm s}^{-1}$ . This assumes that the dominant source of error is the barotropic component deduced from the LADCP, and that the errors are horizontally uncorrelated. We determine the net inflow of water denser than  $\gamma^n = 28.31 \text{ kg m}^{-3}$  and the associated mean volume-transport-weighted temperature and salinity of this water from the direct ALBATROSS measurements. The volume transport of  $4 \text{ Sv}$  is consistent with heat and mass budgets of the Weddell gyre<sup>23</sup>.

### Outflow area, outflow properties and property gradients

The outflow potential temperature and salinity across the  $\gamma^n = 28.31 \text{ kg m}^{-3}$  isopycnal, the area of this isopycnal surface, and the vertical gradients of potential temperature and salinity in the WSDW layer in the Scotia Sea, were calculated using climatological deep hydrographic stations. The Hydrographic Atlas of the Southern Ocean (HASO) database<sup>30</sup> was used as the primary source of quality controlled hydrographic data within the Scotia Sea basin. ALBATROSS stations and the 1993 crossing of the World Ocean Circulation Experiment (WOCE) Drake passage repeat section SR1b<sup>24</sup> were added at the standard HASO vertical grid spacing, typically  $500 \text{ m}$  at the depths of the  $\gamma^n = 28.31 \text{ kg m}^{-3}$  isopycnal. Comparison with a detailed bathymetric data set<sup>16</sup> excluded all stations that did not extend to within  $250 \text{ m}$  of the sea bed or were in water depths shallower than  $1,000 \text{ m}$ .

To determine the area of the  $\gamma^n = 28.31 \text{ kg m}^{-3}$  isopycnal, we selected all stations with bottom neutral density greater than or equal to  $28.26 \text{ kg m}^{-3}$ , thus including all types of WSDW. This yields 299 stations. The bottom neutral densities were gridded on a  $0.5^\circ \times 0.5^\circ$  grid, with a search radius of  $200 \text{ km}$  and smoothing the influence of each point over  $50 \text{ km}$  (Fig. 2). Experimentation with various gridding options gave an estimate of the uncertainty of  $1.0 \times 10^{11} \text{ m}^2$ , including consideration of the sparsity of the data.

Eighty-three individual hydrographic stations sampled water as dense or denser than  $\gamma^n = 28.31 \text{ kg m}^{-3}$  in the Scotia Sea (Fig. 2). The potential temperature and salinity on this isopycnal were determined for each station, and the median value selected to minimize the effect of outliers. The outflow potential temperature was  $-0.25^\circ \text{C}$ , as we would expect, as the  $-0.25^\circ \text{C}$  isotherm is often chosen as the indicator of the boundary of LWSDW. We allocate this a small uncertainty ( $\pm 0.003^\circ \text{C}$ ); in the calculation of diffusivity the pertinent quantity is the difference between the inflow and outflow temperatures,  $\theta_i$  minus  $\theta_o$ , to which we allocate a combined uncertainty of  $\pm 0.01^\circ \text{C}$ . For comparison, the salinity difference is  $0.009$  with an estimated uncertainty of  $0.005$ , indicating the greater robustness of the heat budget calculation.

The same stations provided an estimate of the vertical gradient of potential temperature and salinity at the depth of the  $\gamma^n = 28.31 \text{ kg m}^{-3}$  isopycnal. This is one of the least well defined quantities in the heat budget calculation, because of poor vertical and horizontal data density. If we grid the same 83 stations and produce an area-weighted mean, the

difference in the resulting  $\partial\theta/\partial z$  is  $0.2 \times 10^{-4} \text{ }^{\circ}\text{C m}^{-1}$ . Taking this as the uncertainty in  $\partial\theta/\partial z$ , together with the uncertainty caused by the relatively small number of stations, we allocate an uncertainty of  $0.3 \times 10^{-4} \text{ }^{\circ}\text{C m}^{-1}$ . The vertical salinity gradient at this isopycnal was  $6.8 \times 10^{-6} \text{ m}^{-1}$  with an estimated uncertainty of  $0.5 \times 10^{-6} \text{ m}^{-1}$ .

## Geothermal heating

In some regions of the sea floor, geothermal heating is negligible. Flow through the oceanic crust is a function of the age of the crust; younger crust provides greater heating. The Scotia Sea exhibits a wide variety of ages. We estimate a typical value for  $G$  to be  $0.1 \text{ W m}^{-2}$ , and use this in the heat budget equation with an uncertainty of  $0.1 \text{ W m}^{-2}$ .

## Seawater density, specific heat capacity

For the situation where geothermal heating is included, values for the *in situ* density of sea water and the specific heat capacity are required. We used the HASO climatological data set to determine representative median values. The uncertainty in these is negligible compared with other uncertainties in the calculation.

Received 9 October; accepted 19 December 2001.

1. Munk, W. H. Abyssal recipes. *Deep-Sea Res.* **13**, 207–230 (1966).
2. Munk, W. H. & Wunsch, C. Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res.* **145**, 1997–2010 (1998).
3. Toole, J. M. & McDougall, T. J. in *Ocean Circulation and Climate* (eds Siedler, G., Church, J. & Gould, J.) 337–355 (Academic, London, 2001).
4. Toole, J. M., Polzin, K. L. & Schmitt, R. W. Estimates of diapycnal mixing in the abyssal ocean. *Science* **264**, 1120–1123 (1994).
5. Whitehead, J. A. & Worthington, L. V. The flux and mixing rates of Antarctic Bottom Water within the North Atlantic. *J. Geophys. Res.* **87**, 7903–7924 (1982).
6. Hogg, N. G., Biscaye, P., Gardner, W. & Schmitz, W. J. On the transport and modification of Antarctic Bottom water in the Vema Channel. *J. Mar. Res.* **40** (suppl.), 231–263 (1982).
7. Saunders, P. M. Flow through Discovery gap. *J. Phys. Oceanogr.* **17**, 631–643 (1987).
8. Barton, E. D. & Hill, A. E. Abyssal flow through the Amiranse Trench (Western Indian Ocean). *Deep-Sea Res.* **A 36**, 1121–1126 (1989).
9. Durrieu de Madron, X. & Weatherly, G. Circulation, transport and bottom boundary layers of the deep currents in the Brazil Basin. *J. Mar. Res.* **52**, 583–638 (1994).
10. Roemmich, D., Huata, S. & Rudnick, D. Northward abyssal transport through the Samoan Passage and adjacent regions. *J. Geophys. Res.* **101**, 14039–14055 (1996).
11. McCarthy, M. C., Talley, L. D. & Baring, M. O. Deep upwelling and diffusivity in the southern Central Indian Basin. *Geophys. Res. Lett.* **24**, 2801–2804 (1997).
12. Morris, M. Y., Hall, M. M., St Laurent, L. C. & Hogg, N. G. Abyssal mixing in the Brazil Basin. *J. Phys. Oceanogr.* **31**, 3331–3348 (2001).
13. Polzin, K. L., Speer, K. G., Toole, J. M. & Schmitt, R. W. Intense mixing of Antarctic Bottom Water in the equatorial Atlantic Ocean. *Nature* **380**, 54–57 (1996).
14. Polzin, K. L., Toole, J. M., Ledwell, J. R. & Schmitt, R. W. Spatial variability of turbulent mixing in the abyssal ocean. *Science* **276**, 93–96 (1997).
15. Ledwell, J. R. *et al.* Evidence for enhanced mixing over rough topography in the abyssal ocean. *Nature* **403**, 179–182 (2000).
16. Smith, W. H. F. & Sandwell, D. T. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1956–1962 (1997).
17. Huang, R. X. Mixing and energetics of the oceanic thermohaline circulation. *J. Phys. Oceanogr.* **29**, 727–746 (1999).
18. Zhang, J., Schmitt, R. W. & Huang, R. X. The relative influence of diapycnal mixing and hydrologic forcing on the stability of the thermohaline circulation. *J. Phys. Oceanogr.* **29**, 1096–1108 (1999).
19. Ganachaud, A. & Wunsch, C. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* **408**, 453–457 (2000).
20. Webb, D. J. & Suginobara, N. Vertical mixing in the ocean. *Nature* **409**, 37 (2001).
21. Naveira Garabato, A. C., Heywood, K. J. & Stevens, D. P. Modification and pathways of Southern Ocean deep waters in the Scotia Sea. *Deep-Sea Res.* **I** (in the press).
22. Jackett, D. R. & McDougall, T. J. A neutral density variable for the world's oceans. *J. Phys. Oceanogr.* **27**, 237–263 (1997).
23. Naveira Garabato, A. C., McDonagh, E. L., Stevens, D. P., Heywood, K. J. & Sanders, R. J. On the export of Antarctic Bottom Water from the Weddell Sea. *Deep-Sea Res.* **II** (2002).
24. Arhan, M., Heywood, K. J. & King, B. A. The deep waters from the Southern Ocean at the entry to the Argentine Basin. *Deep-Sea Res.* **II 46**, 475–499 (1999).
25. Orsi, A. H., Johnson, G. C. & Bullister, J. L. Circulation, mixing and production of Antarctic Bottom Water. *Prog. Oceanogr.* **43**, 55–109 (1999).
26. Egbert, G. D. & Ray, R. D. Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. *Nature* **405**, 775–778 (2000).
27. Sjöberg, B. & Stigebrandt, A. Computations of the geographical distribution of the energy flux to mixing processes via internal tide and the associated vertical circulation in the ocean. *Deep-Sea Res.* **39**, 269–291 (1992).
28. Bell, T. H. Topographically-generated internal waves in the open ocean. *J. Geophys. Res.* **80**, 320–327 (1975).
29. Fukumori, I., Raghunath, R. & Fu, L.-L. Nature of global large-scale sea level variability in relation to atmospheric forcing: a modelling study. *J. Geophys. Res.* **103**, 5493–5512 (1998).
30. Olbers, D., Gouretski, V., Seiss, G. & Schroter, J. *Hydrographic Atlas of the Southern Ocean* (Alfred Wegener Inst. for Polar and Marine Res., Bremerhaven, Germany, 1992).

## Acknowledgements

We thank P. Barker and F. Vine for discussions regarding the magnitude of geothermal heating in the Scotia Sea. ALBATROSS was funded by the Natural Environment Research Council.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for information should be addressed to K.J.H. (e-mail: k.heywood@uea.ac.uk).

# Sudden aseismic fault slip on the south flank of Kilauea volcano

Peter Cervelli<sup>†</sup>, Paul Segall<sup>\*</sup>, Kaj Johnson<sup>\*</sup>, Michael Lisowski<sup>‡</sup> & Asta Miklius<sup>†</sup>

<sup>\*</sup> Department of Geophysics, Mitchell Building, Stanford University, Stanford, California 94305-2215, USA

<sup>†</sup> US Geological Survey, Hawaiian Volcano Observatory, Hawaii National Park, Hawaii 96718, USA

<sup>‡</sup> US Geological Survey, Cascades Volcano Observatory, 5400 MacArthur Blvd, Vancouver, Washington 98661, USA

One of the greatest hazards associated with oceanic volcanoes is not volcanic in nature, but lies with the potential for catastrophic flank failure<sup>1,2</sup>. Such flank failure can result in devastating tsunamis and threaten not only the immediate vicinity, but coastal cities along the entire rim of an ocean basin<sup>3</sup>. Kilauea volcano on the island of Hawaii, USA, is a potential source of such flank failures<sup>3,4</sup> and has therefore been monitored by a network of continuously recording geodetic instruments, including global positioning system (GPS) receivers, tilt meters and strain meters. Here we report that, in early November 2000, this network recorded transient southeastward displacements, which we interpret as an episode of aseismic fault slip. The duration of the event was about 36 hours, it had an equivalent moment magnitude of 5.7 and a maximum slip velocity of about 6 cm per day. Inversion of the GPS data reveals a shallow-dipping thrust fault at a depth of 4.5 km that we interpret as the down-dip extension of the Hilina Pali–Holei Pali normal fault system. This demonstrates that continuously recording geodetic networks can detect accelerating slip, potentially leading to warnings of volcanic flank collapse.

Data from the Kilauea GPS network (Fig. 1) are routinely analysed in 24-h batches at the US Geological Survey's Hawaiian Volcano Observatory (HVO). Figure 2 depicts the time series of selected stations on Kilauea's south flank as well as of three more distant stations. The south flank stations show clear offsets of up to 15 mm in early November, while the remote sites do not. The estimated displacements between 8 November and 10 November determined from the daily solutions are shown in Fig. 1. The spatial coherence of the signal and the negligible displacements far from the south flank (Fig. 1) rule out the possibility that the displacements are an artefact of unmodelled errors in the GPS time series.

We interpret the deformation field as resulting from fault slip. The data cannot be fitted by a dyke, sill, or expanding magma chamber. Moreover, there is no evidence of summit deflation, which usually accompanies magmatic activity. Nonlinear optimization<sup>5</sup> of the observed displacements for the best-fitting fault geometry yields a thrust, which dips shallowly ( $4^{\circ} \pm 10^{\circ}$ , uncertainties estimated using a bootstrap method<sup>6</sup>) toward the volcano (Figs 1 and 3). The average slip on the fault is 87 mm. The model-predicted displacements fit the observations, including the vertical components (Figs 1 and 3), extremely well (normalized mean square error of 1.2). The largest south flank earthquake during the first two weeks of November was a magnitude 2.7, which occurred on 3 November (Fig. 1). Thus, this event can be placed among the class of 'silent' earthquakes, which have been recently detected by large-scale

continuously recording geodetic networks<sup>7,8</sup>. This is, however, the first recording of a silent earthquake in a volcanic environment.

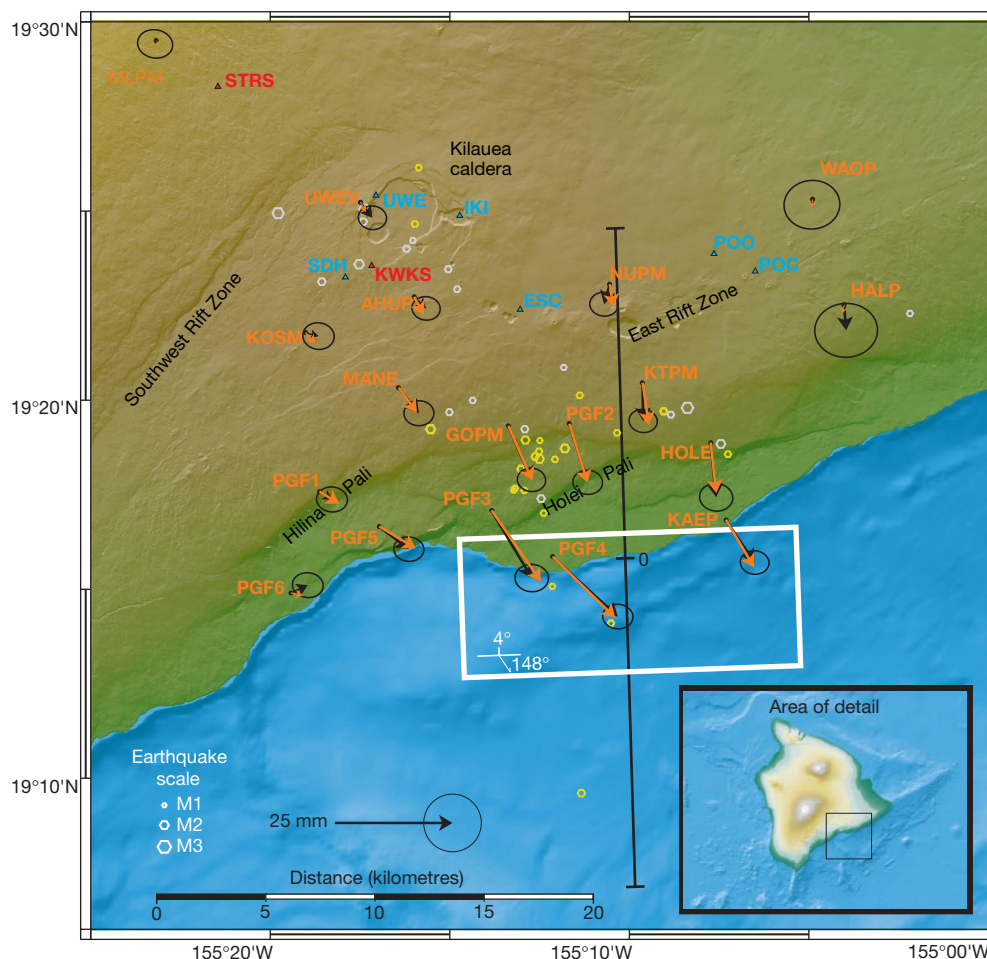
Many authors have suggested that the persistent seaward motion of Kilauea's south flank results from slip along a basal decollement<sup>9–14</sup> at a depth of 7–8 km (Fig. 3). The inferred depth of the November 2000 event, however, is too shallow for it to have occurred on this structure. Indeed, the uncertainties from the bootstrap analysis reject a decollement source at better than 95% confidence. Failure to account for material heterogeneity in the earth is thought to bias geodetic estimates of fault depth<sup>15</sup>. To address this concern, we used the best-fitting fault from the analysis above, which had assumed homogeneity, as a starting point for an inversion in a three-layer earth model. Green's functions for this model were computed using propagator matrices<sup>16,17</sup>. Rigidity was estimated from a seismic velocity model for Kilauea<sup>18</sup>. The estimated source depth does indeed increase, but only by 0.5 km. Assuming the variance of the depth estimated for the homogeneous model is a reasonable approximation to the actual variance, we can still rule out a decollement source with high confidence.

Other possible sources include listric normal faults associated with the Hilina Pali–Holei Pali fault system and shallow thrusts within the south flank wedge (Figs 1 and 3). Seismic reflection profiling<sup>19</sup> has imaged landward dipping reflectors that have been interpreted as thrust faults. Some of these structures have the same

dip and approximate location as the modelled source of the November 2000 slip event (see, for example, Fig. 4 of ref. 19). Taken together, the normal faults and shallow thrusts are thought to comprise a large landslide structure. The data are consistent with the slip event having occurred on such a structure.

The daily GPS solutions put only broad constraints on the duration of the slip event. Conventional kinematic processing of the GPS data to produce solutions at shorter intervals should be capable of resolving a signal of 15 mm (refs 20, 21). However, the stations with the maximum displacements were offline during the slip event because of extensive flooding the previous week. Moreover, of the remaining stations, the one with the maximum displacement (KAEP) is situated in a high multipath environment, making sub-daily solutions poor. For these reasons, kinematic processing failed to demonstrate a convincing signal at individual stations.

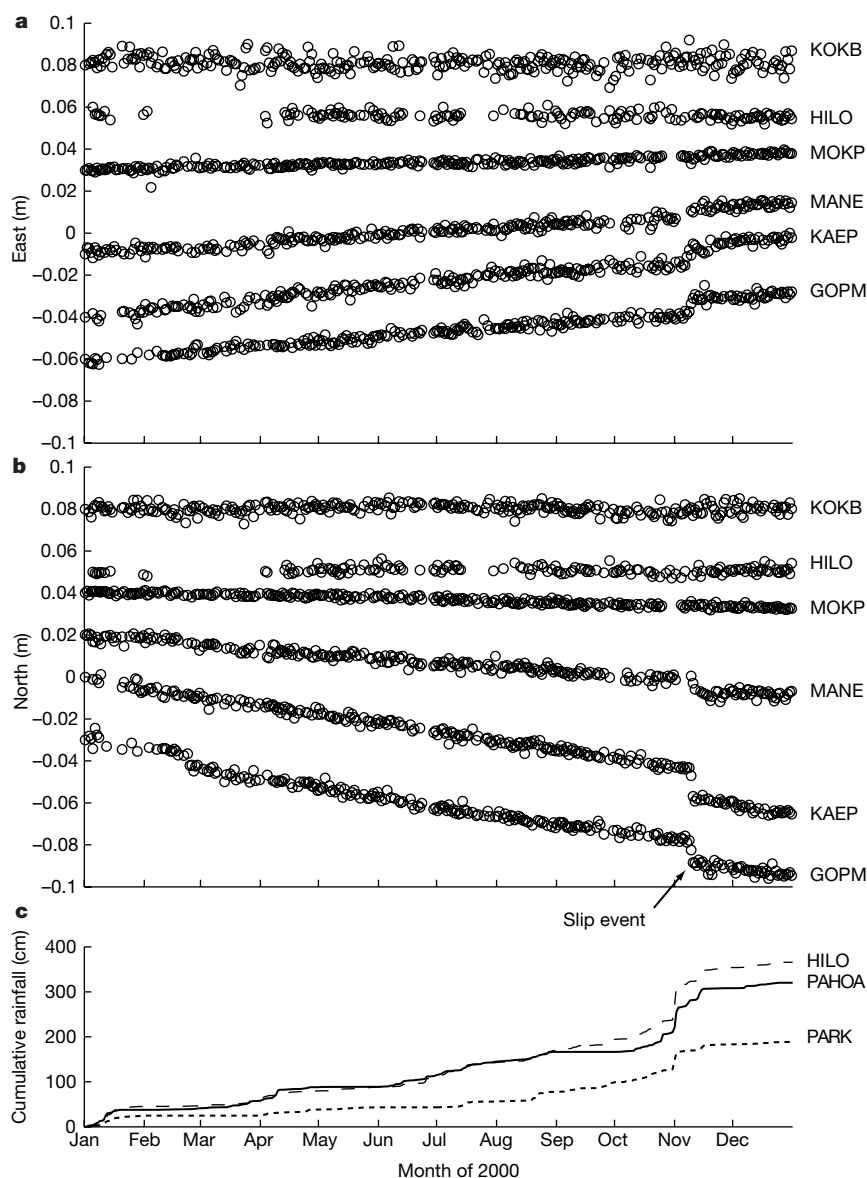
Instead, we used the raw GPS observations to solve directly for fault slip as a function of time, as described in the Methods section. We find that slip began sometime after 18:00 on 8 November 2000 GMT and then proceeded at a nearly steady rate for about 36 hours (Fig. 4). Measurements from a borehole tilt meter (ESC, Fig. 1) and strain meter (STRS, Fig. 1) provide independent support for our estimate of the source duration. Figure 4 compares the observed tilt and strain to that predicted by the time-dependent slip model,



**Figure 1** Observed (black) and predicted (orange) displacements from the November 2000 deformation event on Kilauea volcano. Error ellipses show 95% confidence regions, based on the scaled covariance of the GPS data (scaling was estimated from the day-to-day repeatability). Stations in white are continuous GPS; green stations are tilt meters; red stations are strain meters. Hexagons depict seismicity from 1 to 14 November 2000; those shown in yellow have source times within 36 h of the slip event. The white rectangle

is the surface projection of the best-fitting source model (a uniform slip dislocation). The dimensions of the fault are  $16 \times 7$  km; the total slip was 87 mm with a rake of  $148^\circ$ . These values and a nominal shear modulus of  $3 \times 10^{10}$  Pa lead to an equivalent seismic moment magnitude of M5.7. The black line locates the cross-section shown in Fig. 3. Inset, location of study area on the Big Island of Hawaii.





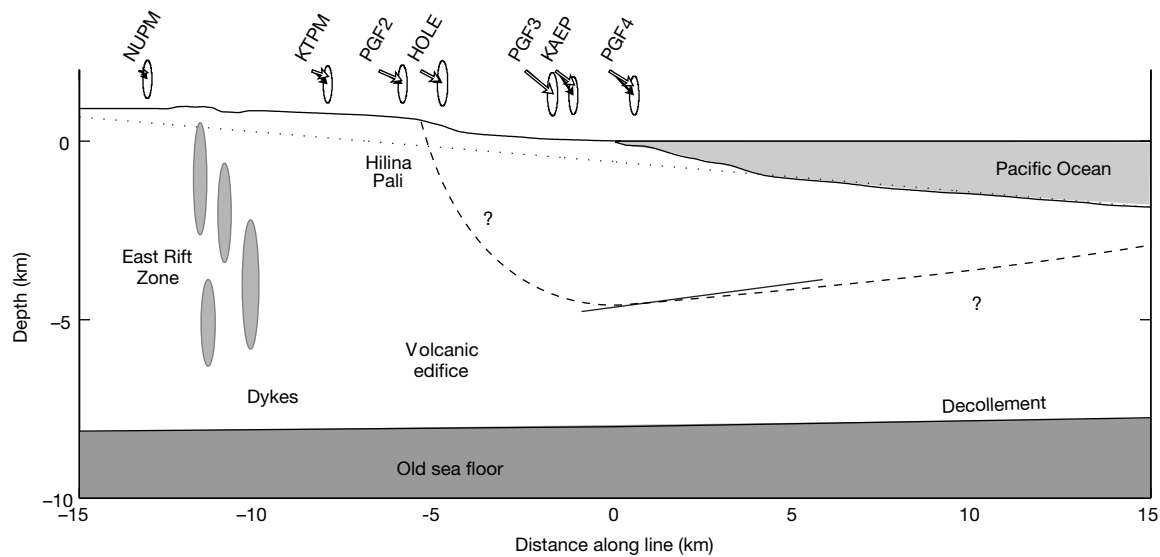
**Figure 2** Horizontal displacement time series of six continuous GPS stations for 2000. The velocity of the station MKEA, near the summit of Mauna Kea, has been removed. **a, b**, The top three time series are from stations distant from Kilauea's south flank. KOKB is in Kokee Park, Kauai, HILO is at the airport in Hilo, Hawaii, and MOKP is on the summit of Mauna

Loa volcano. **a, b**, The bottom three time series are from stations on the south flank (Fig. 1). Note the significant offsets that occur in the south-flank time series that do not occur at the more distant stations. **c**, Rainfall for the year 2000, measured at Hilo Airport, Pahoa School and Hawaii National Park Headquarters.

which was estimated using GPS data alone. Data from the remaining tilt and strain meters are substantially noisier, and although they are consistent with the model predictions, provide no additional information. The lack of a signal at the broad-band seismometer at Kipapa (on the Hawaiian island Oahu) is again consistent with the inferred source duration.

Nine days before the slip event, nearly 1 m of rain fell over the southeastern Big Island (Fig. 2). The timing of this storm suggests a possible causal relationship between the rainfall and the aseismic slip. Rainfall has two effects on fault stability. First, the surface loading instantaneously perturbs the stress throughout the south flank. Second, the elevated water table causes a pore pressure increase that diffuses into the volcano, decreasing the effective normal stress on faults and thereby bringing them closer to failure. The loading probably increased the Coulomb stress at the fault, but the magnitude of the change is a small fraction of the total surface load of approximately 0.1 MPa (for 1 m of water). The change in pore pressure induced at shallow depths is given by:  $\Delta P = \rho g \Delta H / \phi$ ,

where  $\rho$  is the density of water,  $g$  is the gravitational acceleration,  $\Delta H$  is the change in water-table height and  $\phi$  is porosity. A 1-m water-table change and a porosity of  $\phi = 0.05$  increases pore pressures by 2 MPa, an amount capable of triggering fault slip<sup>22</sup>. For this pressure change to have triggered the slip, the fault-zone permeability must be sufficiently high for the pressure front to have reached the source region within the 10 days or so separating the rainfall and slip events. This requires a fault-zone diffusivity of  $c = L^2/4t \approx 30 \text{ m}^2 \text{ s}^{-1}$ , where  $L \approx 10^4 \text{ m}$  and  $t \approx 10$  days. The corresponding permeability is given by  $k = \eta \phi \beta c$ , where  $\eta$  is the viscosity of water and  $\beta$  is a combination of pore and water compressibility. Using  $\beta = 10^{-8} \text{ Pa}^{-1}$ , a value for crack-like pores<sup>23</sup>, we estimate an average permeability of  $k \approx 10^{-10} \text{ m}^2$ , which is consistent with permeability estimates<sup>24</sup> for highly fractured fault zones in Kilauea's south flank. A large fraction of the pressure change could have reached the source region if the permeability of the fault zone significantly exceeds that of the surrounding rock, preventing the pore pressure from dissipating into the



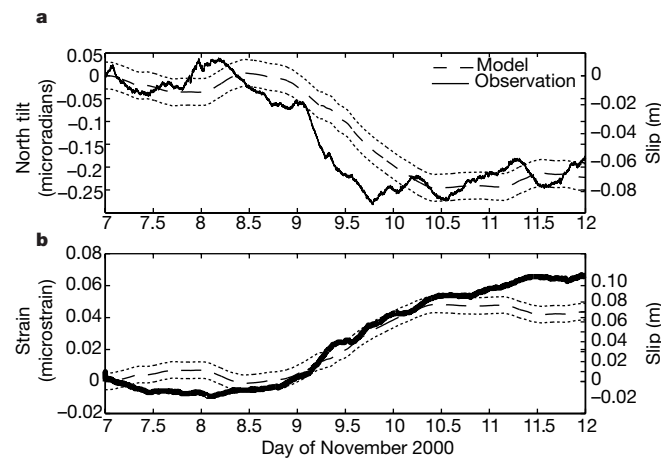
**Figure 3** A cross-section of Kilauea's south flank along the grey line in Fig. 1. The geologic structure is inferred from seismic reflection data<sup>19</sup>. The location of the basal decollement is shown as a solid black line<sup>11</sup>. The dashed black line represents the top of the elastic body used as our earth model. This surface differs from the local horizontal because the elevation decreases monotonically to the southeast. In the source inversion, the dip and depth of the model fault are measured with respect to the top of the half-

space. The model fault (solid line) lies near a region of thrust faults inferred from seismic reflection data<sup>19</sup>, and may connect with the Hilina Pali–Holei Pali normal fault system (dashed). The cross-sectional area of the wedge above the fault system is about 50 km<sup>2</sup>; extending the wedge 20 km in each direction along the strike give a potential slide volume of about 2,000 km<sup>3</sup>, in the event of total disaggregation and failure.

volcano. Given the 10<sup>-15</sup> m<sup>2</sup> permeability typical of the ash layers and the gabbroic rifts that surround the fault zone, this condition may have been met, although permeabilities at depth are highly uncertain.

From a geological perspective, the style and location of deformation from this event supports the view that the large normal faults cutting Kilauea's south flank are relatively shallow features that do not root into the decollement<sup>25–27</sup>. From a hazards perspective, flank

failure on oceanic volcanoes represents an uncommon but significant danger to coastal cities. As these flanks begin to fail, they must accelerate. If slow acceleration precedes catastrophic failure, we have demonstrated that a well-placed geodetic network can detect such an acceleration, even when the absolute rates are small. The rate of slip observed in early November 2000 was about 5 cm per day, far faster than the slip rates (decimetres per year) inferred for Kilauea's decollement, but substantially less than catastrophic failure. For this slip event, the detection was after the fact, but there is no reason why the geodetic data could not be analysed in real time to provide a warning of sudden acceleration and potential flank instability. □



**Figure 4** The source time function for November 2000 deformation event. **a**, The magnitude of the slip on the model fault (right scale, dashed line) and 95% uncertainty regions (dotted line), the predicted north tilt at station ESC (left scale, dashed line) and the observed north tilt at ESC (left scale, solid line), all as functions of time. **b**, Slip magnitude (right scale, dashed line) and 95% uncertainty regions (dotted line), predicted horizontal volumetric strain at STRS (left scale, dashed line) and observed volumetric strain at STRS (left scale, solid line). The apparent time lag between the observed and predicted tilt may be due either to the simplicity of the model (that is, the source geometry was fixed and slip was uniform) or to correlated errors in the tilt record. We varied the scale of the random walk process,  $\sigma_s$ , by a factor of 4 in each direction and found that the estimated duration of the event did not change significantly.

## Methods

### Estimating net displacements

We estimated the net displacements of the slip event from the daily GPS solutions, which are processed at HVO with the Gipsy/Oasis II software package<sup>28</sup> (<http://gipsy.jpl.nasa.gov>). We use a Kalman filter to model time variation of the station positions as the sum of the nominal station position, the long-term station velocity multiplied by the time, and a linearized Helmert transformation multiplied by a vector of reference frame parameters. This reduces the effect of reference frame uncertainty and allows empirical scaling of the GPS data covariance to a realistic value. The frame terms are assumed to be uncorrelated in time. The variance of the nominal station position term is reset to a large value at several specific epochs that correspond to dyke intrusions and earthquakes. This allows geologic offsets to exist in the time series without biasing the frame and the long-term velocity estimates.

### Estimating the source time function

We used the raw GPS observations, which are sampled every 30 s, to solve directly for fault slip using the homogeneous elastic Green's functions from the best-fitting fault. This takes advantage of the spatial coherence of the deformation signal in the GPS network, effectively increasing the signal-to-noise ratio in the estimation of the temporal evolution of the source. Specifically, we model the double-differenced ionosphere-free carrier phase<sup>29</sup>,  $\Phi(t)$  (with wavelength  $\lambda$ ) as:

$$\Phi(t) - r_0 = A\delta\mathbf{x}(t) + \mathbf{z}(t)m_w + \lambda N \quad (1)$$

where  $r_0$  are the predicted double-differenced ranges from the stations to the satellites,  $A$  is a matrix of partial derivatives of the double-difference ranges with respect to station coordinates,  $\delta\mathbf{x}$  is a vector of time-varying differences in station coordinates from their nominal values,  $\mathbf{z}$  is the time-varying tropospheric zenith delay,  $m_w$  is a tropospheric mapping function and  $N$  are the sums of the scaled integer phase ambiguities at the two GPS frequencies. The phase ambiguities suffer occasional cycle slips, which we detect and

flag in advance;  $N$  is reset when such slips occur. The double-differenced pseudoranges are modelled in the same way, with the exception of the phase ambiguity term, which is not needed. Instead of solving for  $\delta x$  for all the stations independently, we solve for a single slip parameter  $s(t)$ , such that  $\delta x(t) = Gs(t)$ , where  $G$  are elastic functions relating slip on the model fault to displacement. We allow the slip to vary as a random walk in time, with scale  $\sigma_s$ . A nominal value of  $\sigma_s = 3 \text{ mm day}^{-1/2}$  is small enough to limit the scatter during periods without anomalous deformation, but large enough to allow the full 87 mm of slip to accumulate over approximately 24 hours. The time-dependent model fits the GPS double-difference phases to within 4 mm (root mean square).

Received 30 August 2001; accepted 21 January 2002.

- McGuire, W. J. in *Volcano Instability on the Earth and Other Planets* (eds McGuire, W. J., Jones, A. P. & Neuberg, J.) 1–23 (Geological Society Spec. Publ. No. 110, London, 1996).
- Moore, J. G. et al. Prodigious submarine landslides on the Hawaiian Ridge. *J. Geophys. Res.* **94**, 17465–17484 (1989).
- Ward, S. Landslide tsunami. *J. Geophys. Res.* **106**, 11201–11216 (2001).
- Lipman, P. W., Normark, W. R., Moore, J. G., Wilson, J. B. & Gutmacher, C. E. The giant submarine Alaka debris slide, Mauna Loa, Hawaii. *J. Geophys. Res.* **94**, 4279–4299 (1988).
- Cervelli, P., Murray, M. H., Segall, P., Aoki, Y. & Kato, T. Estimating source parameters from deformation data, with an application to the March 1997 earthquake swarm off the Izu Peninsula, Japan. *J. Geophys. Res.* **106**, 11217–11238 (2001).
- Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).
- Linde, A. T., Gladwin, M. T., Johnston, M. J. S., Gwyther, R. L. & Bilham, R. G. A slow earthquake sequence on the San Andreas fault. *Nature* **383**, 65–68 (1996).
- Dragert, H., Wang, K. & James, T. S. A silent slip event on the deeper Cascadia subduction interface. *Science* **292**, 1525–1528 (2001).
- Nakamura, K. Why do long rift zones develop in Hawaiian volcanoes—A possible role of thick oceanic sediments. *Bull. Volcanol. Soc. Jpn* **25**, 255–267 (1980) (in Japanese).
- Dieterich, J. H. Growth and persistence of Hawaiian rift zones. *J. Geophys. Res.* **93**, 4258–4270 (1988).
- Thurber, C. H. & Gripp, A. E. Flexure and seismicity beneath the south flank of Kilauea Volcano and tectonic implications. *J. Geophys. Res.* **93**, 4271–4278 (1988).
- Delaney, P. T., Miklius, A., Árnadóttir, T., Okamura, A. T. & Sako, M. K. Motion of Kilauea Volcano during sustained eruption from the Pu'u O'o and Kupaianaha vents, 1983–1991. *J. Geophys. Res.* **98**, 17801–17820 (1993).
- Owen, S. et al. Rapid deformation of the south flank of Kilauea Volcano, Hawaii. *Science* **267**, 1328–1332 (1995).
- Owen, S. et al. Rapid deformation of Kilauea volcano: GPS measurements between 1990 and 1996. *J. Geophys. Res.* **105**, 18983–18998 (2000).
- Du, Y., Segall, P. & Gao, H. Quasi-static dislocations in three dimensional inhomogeneous media. *Geophys. Res. Lett.* **24**, 2347–2350 (1997).
- Aki, K. & Richards, P. G. *Quantitative Seismology: Theory and Methods* (Freeman, San Francisco, 1970).
- Ward, S. N. Quasi-static propagator matrices: creep on strike-slip faults. *Tectonophysics* **120**, 83–106 (1985).
- Okubo, P. G., Benz, H. M. & Chouet, B. A. Imaging the crustal magma sources beneath Mauna Loa and Kilauea volcanoes, Hawaii. *Geology* **25**, 867–870 (1997).
- Morgan, J. K., Moore, G. F., Hills, D. J. & Leslie, S. Overthrusting and sediment accretion along Kilauea's mobile south flank, Hawaii: Evidence for volcanic spreading from marine seismic reflection data. *Geology* **28**, 667–670 (2000).
- Elosegui, P., Davis, J., Johansson, J. & Shapiro, I. Detection of transient motions with the Global Positioning System. *J. Geophys. Res.* **101**, 11249–11261 (1996).
- Larson, K. et al. Volcano monitoring using kinematic GPS: Filtering strategies. *J. Geophys. Res.* **106**, 19453–19464 (2001).
- Stein, R. S. The role of stress transfer in earthquake occurrence. *Nature* **402**, 605–609 (1999).
- David, C., Wong, T.-F., Zhu, W. & Zhang, J. Laboratory measurements of compaction-induced permeability change in porous rocks; implications for the generation and maintenance of pore pressure excess in the crust. *Pure Appl. Geophys.* **143**, 425–456 (1994).
- Ingebritsen, S. E. & Scholl, M. A. The hydrology of Kilauea Volcano. *Geothermics* **22**, 255–270 (1993).
- Swanson, D. A., Duffield, W. A. & Fiske, R. S. Displacement of the south flank of Kilauea Volcano: The result of forceful intrusion of magma into the rift zones. *US Geol. Surv. Prof. Pap.* **963**, 1–30 (1976).
- Dvorak, J. J., Klein, F. W. & Swanson, D. A. Relaxation of the south flank after the 7.2-Magnitude Kalapana earthquake, Kilauea volcano, Hawaii. *Bull. Seismol. Soc. Am.* **84**, 133–141 (1994).
- Cannon, E. C., Bürgmann, R. & Owen, S. E. Shallow normal faulting and block rotation associated with the 1975 Kalapana earthquake, Kilauea Volcano, Hawaii. *Bull. Seismol. Soc. Am.* (in the press).
- Lichten, S. & Borden, J. Strategies for high-precision Global Positioning System orbit determination. *J. Geophys. Res.* **92**, 12751–12762 (1987).
- Hofmann-Wellenhof, B., Lichtenegger, H. & Collins, J. *GPS: Theory and Practice* (Springer, New York, 1994).

## Acknowledgements

We thank M. Bevis for access to the GPS data for the University of Hawaii network, P. Okubo for providing hypocentres, G. Blewitt, R. Bürgmann, J. Davis, R. Iversen, M. Johnston, J. McGuire, J. Murray, D. Swanson and W. Thatcher for comments and discussion, and D. Okita for getting us to and from Kilauea GPS stations safely and efficiently. The National Science Foundation and the US Geological Survey provided funding for this research.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C. (e-mail: cervelli@sumo.wr.usgs.gov).

# Tyrannosaurus was not a fast runner

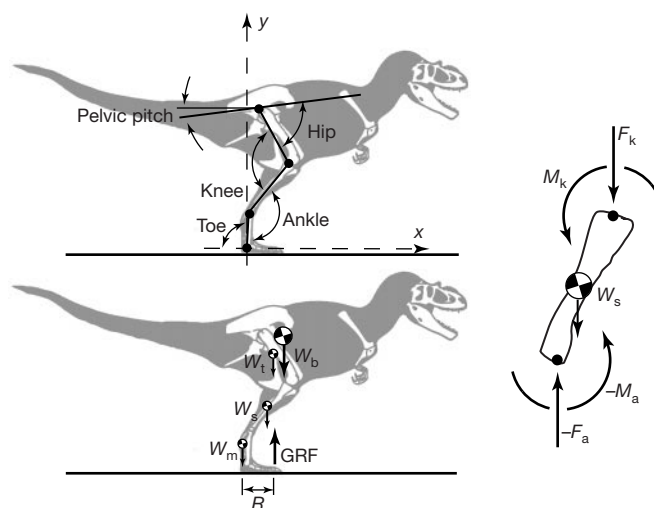
John R. Hutchinson\*† & Mariano Garcia\*†

\* Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA

† Present addresses: Biomechanical Engineering Division, Stanford University, Stanford, California 94305-4038, USA (J.R.H.); Borg-Warner Automotive, 770 Warren Road, Ithaca, New York 14850, USA (M.G.)

The fastest gait and speed of the largest theropod (carnivorous) dinosaurs, such as *Tyrannosaurus*, is controversial. Some studies contend that *Tyrannosaurus* was limited to walking, or at best an  $11 \text{ m s}^{-1}$  top speed<sup>1–4</sup>, whereas others argue for at least  $20 \text{ m s}^{-1}$  running speeds<sup>5–7</sup>. We demonstrate a method of gauging running ability by estimating the minimum mass of extensor (supportive) muscle needed for fast running. The model's predictions are validated for living alligators and chickens. Applying the method to small dinosaurs corroborates other studies by showing that they could have been competent runners. However, models show that in order to run quickly, an adult *Tyrannosaurus* would have needed an unreasonably large mass of extensor muscle, even with generous assumptions. Therefore, it is doubtful that *Tyrannosaurus* and other huge dinosaurs (~6,000 kg) were capable runners or could reach high speeds.

Most assessments of running ability in large theropods are qualitative, based on analogies to walking elephants or running birds and hoofed mammals<sup>3,5–7</sup>. Earlier quantitative assessments of the biomechanics of *Tyrannosaurus* suggested that it had limited locomotor performance<sup>1,2,4,8,9</sup>, but uncertainties about tyrannosaur anatomy, physiology, and behaviour have impeded resolution of this debate<sup>1–9</sup>. Fossilized footprints demonstrate that smaller theropod dinosaurs could run<sup>10</sup>, but running tracks from large theropods are unknown<sup>7,9</sup>. Bulky limbs and other analogies with elephants



**Figure 1** Explanation of free-body diagram analysis of body segments. Skeletal illustration modified from ref. 6. The initial *Tyrannosaurus* model (Trex\_1) is shown in right lateral view and ( $x, y$ -coordinate space, with the origin located at the toe joint. Pelvic pitch, hip, knee, ankle and toe joint angle definitions are shown (see Supplementary Information). One of the angles is redundant (four angles suffice). The ground reaction force (GRF) is vertical (typical for mid-stance<sup>12–14</sup>) and acts at a distance  $R$  from the toe joints. The body segment weights ( $W_b$ ,  $W_t$ ,  $W_s$  and  $W_m$  for the trunk, thigh, shank and metatarsus) are also shown. A free-body diagram was used to calculate the internal moments about each joint. For example, about the knee joint (see inset), the moment that knee extensor muscles must generate ( $M_k$ ) is equal to the ankle contact force ( $-F_a$ ) times its respective moment arm, plus the gravitational moment of the shank segment and the ankle moment ( $-M_a$ ).



provide evidence that the enormous sauropod dinosaurs did not run<sup>1,2,5,11</sup>.

For extant animals, our results (Table 1) matched reality: alligators cannot run bipedally, whereas chickens are adept runners. Our dissected *Alligator* only had about 3.6%  $m_{\text{body}}$  per leg as extensor muscles, but we estimated that it would need about 7.7%  $m_{\text{body}}$  per leg to run quickly on its hindlimbs. In contrast, our *Gallus* specimen had about 8.8%  $m_{\text{body}}$  per leg as extensor muscles, but it should only need about 4.7%  $m_{\text{body}}$  per leg for fast running. Thus alligators have less than half the extensor muscle they would need to run fast, whereas chickens have almost twice the required amount of extensor muscle.

Well known scaling principles<sup>12–14</sup> predict that animals of larger body mass ( $m_{\text{body}}$ ) have more restricted locomotor performance because muscle force scales as  $(m_{\text{body}})^{0.67}$ , whereas muscle mass scales as  $(m_{\text{body}})^{1.0}$ . To be a fast runner, a *Tyrannosaurus* would have needed enough extensor muscle mass to support itself. We calculated (Fig. 1; Methods) this mass for various animals to determine whether running would require too much muscle mass.

Smaller theropods had lower estimates of minimum extensor muscle mass per leg ( $T$ ) than *Tyrannosaurus* (Table 1), and thus we expect they had a broader range of running performance. We initially calculated (Fig. 2) that *Tyrannosaurus* needed about 43% of its body mass as extensor muscles in each leg (86%  $m_{\text{body}}$  for both

legs) in order to run quickly. This high estimate of  $T$  is surprising because our input parameter values were generously weighted towards a low  $T$ . These findings suggest that *Tyrannosaurus* could not run quickly.

A parameter study was then performed to assess the sensitivity of these findings to the values of the assumed parameters. Many of the parameters needed to model the limb mechanics of extinct dinosaurs are uncertain. To test how  $T$  was affected by the initial parameters, we entered a range of values and observed how  $T$  changed (Table 2). Some unknown parameters, such as the muscle fibre pennation angle ( $\theta$ ), were relatively unimportant: within a reasonable range of values (compared to extant taxa) they had little or no effect. Likewise, our approximations of the muscle moment arms ( $r$ ) were presumably reliable because we used rigorous muscle reconstructions that minimize speculation<sup>15,16</sup>. Although  $m_{\text{body}}$  is unknown for extinct taxa, our calculation expressed the required extensor muscle mass as a percentage of body mass ( $T$ ), factoring out  $m_{\text{body}}$  (see Methods).

In contrast, uncertainty about other parameter values was quite important. The limb orientation used (Figs 1 and 2) was a critical assumption because a more columnar (straight-legged) orientation (Table 2) decreased the moment arms ( $R$ ) of the ground reaction force (GRF) and hence decreased  $T$  compared to a more crouched (bent-legged) limb orientation. A more posterior position of the trunk centre of mass, shorter extensor muscle fibre lengths ( $L$ ), greater extensor muscle moment arms ( $r$ ), or a lower multiplier ( $G$ ) of the GRF would also decrease our estimated values of  $T$  (Table 2). The relative lengths of muscle fibres ( $L$ ) vary widely in extant taxa<sup>17,18</sup>, so this parameter deserves more study. Admittedly, our calculations omitted other parameters that a more realistic model could include (Supplementary Information). However, most of these assumptions were conservative, leading to an underestimate of  $T$ .

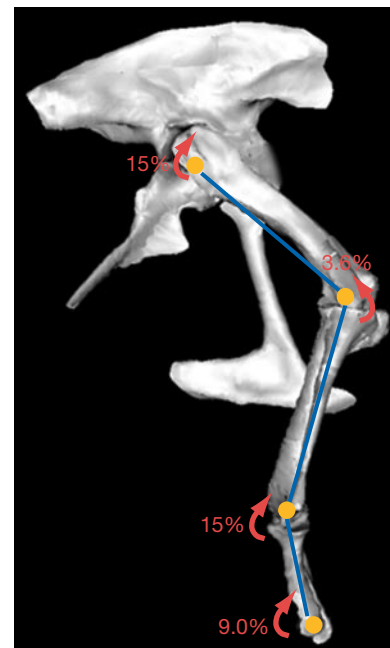
We asked how much hindlimb muscle mass would be unreasonable for an extinct theropod dinosaur: in extant tetrapods, 50% or less of  $m_{\text{body}}$  is typically composed of muscle<sup>19</sup>. This proportion is

**Table 1 Muscle anatomical data entered in equation (1) to estimate  $T$**

|                                          | Hip    | Knee    | Ankle  | Toe    |
|------------------------------------------|--------|---------|--------|--------|
| <i>Alligator</i>                         |        |         |        |        |
| $M_{\text{muscle}}$ (N m)                | 16     | 12      | 12     | 4.7    |
| $L$ (m)                                  | 0.092  | —       | 0.037  | 0.0020 |
| $\theta$ (°)                             | 18     | —       | 21     | 19     |
| $r$ (m)                                  | 0.019  | —       | 0.017  | 0.0050 |
| $m_i$ (% $m_{\text{body}}$ )             | 4.9    | 0       | 1.7    | 1.1    |
| $T$ (% $m_{\text{body}}$ )               |        |         |        | 7.7    |
| <i>Gallus</i>                            |        |         |        |        |
| $M_{\text{muscle}}$ (N m)                | 4.0    | −0.30   | 4.7    | 2.2    |
| $L$ (m)                                  | 0.085  | 0.051   | 0.026  | 0.026  |
| $\theta$ (°)                             | 0      | 25      | 23     | 22     |
| $r$ (m)                                  | 0.038  | 0.026   | 0.010  | 0.0050 |
| $m_i$ (% $m_{\text{body}}$ )             | 0.08   | 2.0     | 2.0    |        |
| $T$ (% $m_{\text{body}}$ )               |        |         |        | 4.7    |
| <i>Coelophysis</i>                       |        |         |        |        |
| $M_{\text{muscle}}$ (N m)                | 45     | −14     | 56     | 44     |
| $L$ (m)                                  | 0.17   | 0.097   | 0.065  | 0.037  |
| $\theta$ (°)                             | 9.0    | 21      | 22     | 21     |
| $r$ (m)                                  | 0.080  | 0.028   | 0.014  | 0.0060 |
| $m_i$ (% $m_{\text{body}}$ )             | 1.4    | 0.73    | 4.0    | 4.1    |
| $T$ (% $m_{\text{body}}$ )               |        |         |        | 10     |
| Small tyrannosaur                        |        |         |        |        |
| $M_{\text{muscle}}$ (N m)                | 570    | −180    | 580    | 390    |
| $L$ (m)                                  | 0.40   | 0.19    | 0.18   | 0.062  |
| $\theta$ (°)                             | 9.0    | 21      | 22     | 21     |
| $r$ (m)                                  | 0.14   | 0.057   | 0.025  | 0.020  |
| $m_i$ (% $m_{\text{body}}$ )             | 4.2    | 1.6     | 11.4   | 3.3    |
| $T$ (% $m_{\text{body}}$ )               |        |         |        | 21     |
| <i>Tyrannosaurus</i> (Trex_1 best guess) |        |         |        |        |
| $M_{\text{muscle}}$ (N m)                | 75,000 | −24,000 | 66,000 | 49,000 |
| $L$ (m)                                  | 1.2    | 0.52    | 0.39   | 0.19   |
| $\theta$ (°)                             | 9.0    | 21      | 22     | 21     |
| $r$ (m)                                  | 0.37   | 0.22    | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ )             | 15     | 3.6     | 15     | 9.0    |
| $T$ (% $m_{\text{body}}$ )               |        |         |        | 43     |

Values given are for all four limb joints (hip, knee, ankle, toe).  $M_{\text{muscle}}$  is the muscle moment required to maintain static equilibrium during single limb support at mid-stance of fast running (with  $G = 2.5$ );  $L$  is mean muscle fibre length;  $\theta$  is mean extensor muscle fibre pennation angle;  $r$  is mean extensor muscle moment arm (means were weighted by physiological cross-sectional area as in refs 12–14);  $m_i$  is the percentage of  $m_{\text{body}}$  needed as extensor muscles about one joint;  $T$  is the total percentage of  $m_{\text{body}}$  per leg needed as extensor muscle (sum of all four  $m_i$  values).

The  $M_{\text{muscle}}$  shown is an extensor moment if it is positive, flexor if negative; except at the knee where the converse is true. In the *Alligator* model, the knee joint  $m_i$  was set at 0 because the knee  $M_{\text{muscle}}$  was a flexor moment and we conservatively assumed that two-joint hip extensors and knee flexors (for example *M. caudofemoralis longus*) could stabilize both joints. See Supplementary Information for more details.



**Figure 2** Limb orientation used for model Trex\_1. The values of  $m_i$ , the percentage of body mass as extensor muscle acting about each joint (muscle moment action shown as a red arrow) required to maintain static equilibrium at mid-stance during fast running are shown in red. Actual hindlimb bones were digitized from *Tyrannosaurus* specimen MOR 555 (Museum of the Rockies; Bozeman, Montana).

independent of  $m_{\text{body}}$  (ref. 20). Of that 50%, a total of 5–40% of  $m_{\text{body}}$  is allocated to limb extensor muscles<sup>15,17–19</sup>. Considering these data, the simplest inference is that if  $T$  is greater than 25%  $m_{\text{body}}$  per leg as limb extensors (that is, more than 50%  $m_{\text{body}}$ ), a biped could not run quickly (if at all), because it would not have enough muscle to exert the necessary forces.

Judging from extant taxa, values of  $T$  from 5–25% per leg indicate that an animal may or may not be a fast runner, depending on the ratio of the actual percentage of  $m_{\text{body}}$  that is extensor muscle to  $T$ , the estimated percentage of extensor muscle required. Values of  $T$  closer to 25% probably indicate poor running ability (if any) because the ratio of actual extensor muscle percentage to  $T$  would be perilously close to 1. Adept runners (such as ground birds<sup>12,21</sup>) have  $T$  values below 5%, maintaining a high ratio of actual extensor muscle mass to  $T$  (almost 2.0 in *Gallus*). This allows the musculoskeletal system to operate with a generous margin of safety<sup>12–14</sup> to accommodate unexpected increases of joint moments.

Some robust conclusions are possible despite the uncertainties about parameter values. Our findings agree with data from fossilized footprints (fastest estimated speeds are  $\sim 11 \text{ m s}^{-1}$ )<sup>9,10</sup> indicating that smaller theropods could run quickly. Although our  $T$

estimates are somewhat high (10–21%  $m_{\text{body}}$  per leg), a more columnar limb orientation could easily have reduced  $T$  enough to enable fast running ( $\sim 5\%$   $m_{\text{body}}$ ).

Yet our estimates of  $T$  for an adult *Tyrannosaurus* were so high that it is difficult to justify the reconstruction of *Tyrannosaurus* as a fast runner. Even with generous assumptions we were unable to reduce the estimates of  $T$  to 5%  $m_{\text{body}}$  per leg or less. Indeed, scaling data<sup>17,18</sup> suggest that a 6,000 kg *Tyrannosaurus* had only 7–10%  $m_{\text{body}}$  per leg as extensors. Therefore we conclude that an adult *Tyrannosaurus* had very limited, if any, running ability. Our best estimates indicate that an adult *Tyrannosaurus* needed over 40%  $m_{\text{body}}$  per leg as extensor muscle. If this is correct, then such animals were unable to run at all. If *Tyrannosaurus* was indeed an adept runner, then it must have had many musculoskeletal specializations that available data do not suggest.

A final model is illustrative. We isometrically scaled-up our chicken model to 6,000 kg to simulate a *Tyrannosaurus*-sized chicken. Our model shows that a gigantic chicken, using the same limb orientation as an extant galliform<sup>21</sup>, would need 99%  $m_{\text{body}}$  per leg as extensor muscles to run quickly, which is clearly impossible and is much higher than for a typical chicken (4.7%  $m_{\text{body}}$  per leg). Features such as the more posterior position of the trunk centre of mass of *Tyrannosaurus* explain why we obtained lower estimates of  $T$  for *Tyrannosaurus* (43%  $m_{\text{body}}$  per leg) than for the 6,000 kg chicken. Nonetheless, the high estimated values of  $T$  in both models reveal how giant terrestrial animals become restricted from extreme locomotor performance<sup>9,13</sup>.

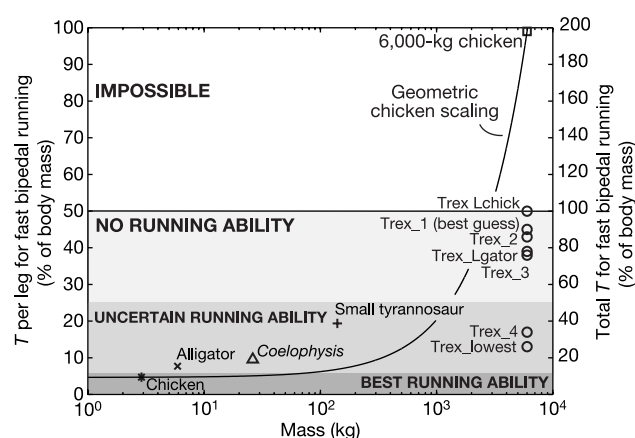
The 'giant chicken' example naturally prompts one to look at equation (1) in terms of scaling. Assuming isometric scaling, and that muscle pennation ( $\theta$ ) and stress ( $\sigma$ ) are independent of body mass<sup>13,17,18,22</sup>, we predict that  $T$  increases linearly with increasing size (Fig. 3). We expect  $T$  to increase with mass roughly in the manner shown by the dashed curve in Fig. 3, especially if limb orientation remains quite flexed.

Our conclusions lead us to question the reconstructions of *Tyrannosaurus* running at 11–20  $\text{m s}^{-1}$  (refs 3–8). Even at a lower running speed of 11  $\text{m s}^{-1}$  (refs 3, 4, 8), an adult *Tyrannosaurus* would have been near or above its maximum muscular capacity. A walking tyrannosaur could have adopted more extended limb joints

**Table 2 Parameter sensitivity in *Tyrannosaurus* models**

|                              | Hip    | Knee    | Ankle  | Toe    |
|------------------------------|--------|---------|--------|--------|
| <b>Trex_2</b>                |        |         |        |        |
| $M_{\text{muscle}}$ (N m)    | 53,000 | –56,000 | 72,000 | 45,000 |
| $L$ (m)                      | 1.2    | 0.52    | 0.39   | 0.19   |
| $\theta$ (°)                 | 9.0    | 21      | 22     | 21     |
| $r$ (m)                      | 0.37   | 0.22    | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ ) | 10     | 8.3     | 16     | 8.3    |
| $T$ (% $m_{\text{body}}$ )   |        |         |        | 43     |
| <b>Trex_3</b>                |        |         |        |        |
| $M_{\text{muscle}}$ (N m)    | 75,000 | 49,000  | 80,000 | 29,000 |
| $L$ (m)                      | 1.2    | –       | 0.39   | 0.19   |
| $\theta$ (°)                 | 9.0    | –       | 22     | 21     |
| $r$ (m)                      | 0.37   | –       | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ ) | 15     | 0       | 18     | 5.3    |
| $T$ (% $m_{\text{body}}$ )   |        |         |        | 38     |
| <b>Trex_4</b>                |        |         |        |        |
| $M_{\text{muscle}}$ (N m)    | 75,000 | 6,500   | 4,400  | 4,100  |
| $L$ (m)                      | 1.2    | –       | 0.39   | 0.19   |
| $\theta$ (°)                 | 9.0    | –       | 22     | 21     |
| $r$ (m)                      | 0.37   | –       | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ ) | 15     | 0       | 0.99   | 0.75   |
| $T$ (% $m_{\text{body}}$ )   |        |         |        | 17     |
| <b>Trex_Lgator</b>           |        |         |        |        |
| $M_{\text{muscle}}$ (N m)    | 75,000 | –24,000 | 66,000 | 49,000 |
| $L$ (m)                      | 0.93   | 0.39    | 0.37   | 0.19   |
| $\theta$ (°)                 | 18     | 16      | 21     | 19     |
| $r$ (m)                      | 0.37   | 0.22    | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ ) | 12     | 2.6     | 14     | 9.0    |
| $T$ (% $m_{\text{body}}$ )   |        |         |        | 38     |
| <b>Trex_Lchick</b>           |        |         |        |        |
| $M_{\text{muscle}}$ (N m)    | 75,000 | –24,000 | 66,000 | 49,000 |
| $L$ (m)                      | 1.1    | 0.65    | 0.41   | 0.33   |
| $\theta$ (°)                 | 0      | 25      | 23     | 22     |
| $r$ (m)                      | 0.37   | 0.22    | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ ) | 13     | 4.6     | 16     | 16     |
| $T$ (% $m_{\text{body}}$ )   |        |         |        | 50     |
| <b>Trex_lowest</b>           |        |         |        |        |
| $M_{\text{muscle}}$ (N m)    | 75,000 | 6,500   | 4,400  | 4,100  |
| $L$ (m)                      | 0.93   | –       | 0.37   | 0.19   |
| $\theta$ (°)                 | 0      | –       | 0      | 0      |
| $r$ (m)                      | 0.37   | –       | 0.11   | 0.065  |
| $m_i$ (% $m_{\text{body}}$ ) | 11     | 0       | 0.87   | 0.71   |
| $T$ (% $m_{\text{body}}$ )   |        |         |        | 13     |

This shows the effect on the estimate of  $T$  (relative to model Trex\_1) when we changed limb orientation (models Trex\_2–4) or muscle-fibre lengths (model Trex\_Lgator) and Trex\_Lchick, isometrically scaling up  $L$  from *Alligator* (5.91 kg) and *Gallus* (2.69 kg) to *Tyrannosaurus* (6,000 kg). As with *Alligator* (Table 1), models Trex\_3, Trex\_4, and Trex\_lowest had a flexor  $M_{\text{muscle}}$  about the knee joint, so the knee  $m_i$  was conservatively set at 0. Model Trex\_lowest shows a combination of parameter values ( $L$  from model Trex\_Lgator, limb orientation from model Trex\_4, and  $\theta$  set at 0°) that brought  $T$  to its lowest estimated value (see Supplementary Information).



**Figure 3** Estimated extensor muscle mass ( $T$ , as a percentage of body mass per leg) required to run quickly. Data are from various models and extant species plotted against  $\log(m_{\text{body}})$  to illustrate the scaling of  $T$ . Larger runners need a larger fraction of their body allocated to extensor muscle mass. Three shades of grey indicate the likelihood of running ability, based on estimated  $T$ . Very low estimated  $T$  implies a good chance of being a capable runner. Very high estimated  $T$  signals little chance of being a capable runner. Intermediate values of estimated  $T$  are inconclusive, and running ability in this case depends on the ratio of actual extensor muscle mass to  $T$ . Values of  $T$  greater than 50% per leg, or greater than 100% in total, are impossible. The solid line represents the case of scaling a chicken isometrically up to 6,000 kg.

to lower the GRF moment arms ( $R$ ) and used double limb support to reduce the GRF, decreasing the  $T$  required for walking within reasonable bounds (Table 2: models *Trex\_4* and *Trex\_lowest*). However, our calculations suggest that even a walking tyrannosaur required activation of a large fraction of its extensor muscle volume, at considerable metabolic expense.

These conclusions should apply to all larger dinosaurs, invalidating arguments that tyrannosaurs were too slow to prey on large contemporaries<sup>3</sup> such as *Triceratops*. *Tyrannosaurus* shows the most extreme cursorial (that is, locomotion-related) specializations of larger theropods<sup>5–9</sup>, but even it was probably a slow runner, at best. Therefore we infer that the numerous lineages of theropod dinosaurs that evolved enormous sizes<sup>9</sup> independently reduced their running ability. Conversely, reduction of body size and relative increases of extensor muscle sizes and moment arms probably increased running ability in maniraptoran theropods, especially birds<sup>15,16</sup>. □

## Methods

The minimum mass of muscle ( $m_i$ ) required to produce the joint moment  $M_{\text{muscle}}$  to maintain static equilibrium about any joint  $i$  is a function of the muscle density ( $d$ ), multiplied by the volume ( $V$ ), divided by the fraction of active muscle volume ( $c$ ).  $V$  is equal to the muscle fibre length ( $L$ ), times the cross-sectional area ( $A$ ), divided by the cosine of the pennation angle of the muscle fibres ( $\cos\theta$ )<sup>12–14</sup>. Furthermore,  $A$  is equal to the muscle force ( $F$ ) divided by the stress ( $\sigma$ ). Finally,  $F$  must be the  $M_{\text{muscle}}$  required to stabilize each joint, divided by the mean extensor muscle moment arm ( $r$ )<sup>12–14</sup>. Combining these expressions to solve for the extensor muscle mass required to act about a particular joint  $i$  ( $m_i$ ; expressed as a percentage of body mass,  $m_{\text{body}}$ ) produces:

$$m_i = (100 M_{\text{muscle}} L d) / (\sigma c r m_{\text{body}} \cos\theta) \quad (1)$$

These parameters do not depend greatly on the physiology of the animal<sup>22</sup>, so the equation holds for both warm- and cold-blooded animals.

## Estimating the joint moment $M_{\text{muscle}}$

In order to estimate  $M_{\text{muscle}}$  (and hence  $m_i$ ), we used MATLAB (The Mathworks) software to compute the joint moments needed to maintain static equilibrium in a particular limb orientation. To maintain equilibrium, muscles must exert net muscle moments about joints that balance equal and opposite moments incurred by external and internal forces, such as the ground reaction force (GRF) and segment weights ( $W_b$ ,  $W_f$ ,  $W_s$  and  $W_m$  in Fig. 1). These moments can be calculated in two dimensions by drawing a set of free-body diagrams and applying the three equilibrium equations ( $\Sigma F_x = 0$ ,  $\Sigma F_y = 0$ ,  $\Sigma M = 0$ ). This procedure is appropriate for standing, and we explain below how it can be applied to running. We summed  $m_i$  for all four limb joints to estimate  $T$ , the total extensor muscle mass required per leg, as a percentage of  $m_{\text{body}}$ .

## Choosing $G$

We multiplied the GRF magnitude in the free-body diagram analysis by a factor  $G$ , the ratio of the GRF to body weight ( $m_{\text{body}}g$ , with  $g = 9.81 \text{ m s}^{-2}$ ), to estimate the  $M_{\text{muscle}}$  needed to support the body on one leg at mid-stance of fast running. At mid-stance of running, the body centre of mass is at its lowest point accelerating upwards, but inertial moments are minuscule relative to the moments of the GRF, so mid-stance is approximated as quasi-static<sup>12–14,23</sup>. The value of  $G$  was set at 2.5, which is a reasonable value for fast running or hopping bipeds<sup>12–14,24,25</sup>.

## Determining forward velocity

To determine an approximate forward velocity that would be consistent with  $G = 2.5$  for *Tyrannosaurus*, we can use the notion of dynamic similarity, such as the Froude number ( $Fr$ ), the ratio of centripetal to gravitational forces<sup>26</sup>, calculated as  $Fr = (\text{velocity})^2 / (\text{hip height} \times g)$ . Our value of  $G = 2.5$  for *Tyrannosaurus* high-speed running is comparable to the value ( $G = 2.7$ ) estimated for an ostrich<sup>27</sup> running at  $12 \text{ m s}^{-1}$  with  $Fr = 16$ . For *Tyrannosaurus* (hip height is roughly 2.5 m),  $Fr = 16$  implies an average forward speed of almost  $20 \text{ m s}^{-1}$ . Therefore our model assumes a GRF magnitude consistent with the faster estimates of *Tyrannosaurus* running speed<sup>3–7</sup>. In theory, at  $Fr > 1$ , an animal should switch from a walk to a run<sup>26</sup>. Consequently the maximum walking speed that an adult tyrannosaur might have used was roughly  $5 \text{ m s}^{-1}$  (11 m.p.h.), which is still faster than speed estimates from known tracks of large theropods<sup>9</sup>.

## Model parameters

To reconstruct running mechanics in extinct dinosaurs, we modelled a small adult theropod (*Coelophysis*), a small juvenile tyrannosaur, and an adult *Tyrannosaurus* (Supplementary Information). Functional segment lengths (distances between centres of joint rotation), masses, and centres of mass were measured from dissections for the extant taxa, or approximated for the extinct taxa. Hip, knee, ankle and toe joints connected these segments and were given joint angles. These parameters (see Supplementary Information) were used in a free-body diagram analysis (Fig. 1) to calculate  $M_{\text{muscle}}$  for equation (1).

We entered other parameter values (Table 1) into equation (1) to estimate  $T$ . The values of  $L$ ,  $\theta$  and  $r$  were measured from dissections for the extant taxa, or approximated for the extinct taxa. The muscle density ( $d$ ) was set at  $1.06 \times 10^3 \text{ kg m}^{-3}$  as in typical skeletal muscle<sup>28</sup>, and the muscle stress ( $\sigma$ ) was entered as  $3.00 \times 10^5 \text{ N m}^{-2}$ , the maximum isometric stress frequently measured<sup>22</sup>. Because we were estimating the minimum mass of muscle required to produce  $M_{\text{muscle}}$ , we set the fraction of active muscle volume ( $c$ ) at 1. Muscle moment arms ( $r$ ) were estimated for extinct taxa by applying muscle reconstructions<sup>15,16</sup> to skeletal material, at appropriate joint angles.

## T constraints

The ratio of actual extensor muscle mass to  $T$ , the required mass (both terms as a percentage of  $m_{\text{body}}$ ) must be at least 1 for a bipedal animal to be able to run quickly; otherwise the actual muscles would be unable to generate the required forces. For extinct animals, we cannot measure the actual muscle masses, but as the estimated  $T$  increases, good running performance becomes increasingly unlikely. If  $T$  is more than 50%  $m_{\text{body}}$  per leg (100%  $m_{\text{body}}$  total), running at  $Fr = 16$  is certainly impossible.

Received 27 July; accepted 12 December 2001.

- Alexander, R. McN. Mechanics of posture and gait of some large dinosaurs. *Zool. J. Linn. Soc.* **83**, 1–25 (1985).
- Alexander, R. McN. *Dynamics of Dinosaurs and Other Extinct Giants* (Columbia Univ. Press, New York, 1989).
- Hornor, J. R. & Lessem, D. *The Complete T. rex* (Simon & Schuster, New York, 1993).
- Farlow, J. O., Smith, M. B. & Robinson, J. M. Body mass, “strength indicator”, and cursorial potential of *Tyrannosaurus rex*. *J. Vert. Paleont.* **15**, 713–725 (1995).
- Bakker, R. T. *Dinosaur Heresies* (William Morrow, New York, 1986).
- Paul, G. S. *Predatory Dinosaurs of the World* (Simon & Schuster, New York, 1988).
- Paul, G. S. Limb design, function and running performance in ostrich-mimics and tyrannosaurs. *Gaia* **15**, 257–270 (2000).
- Christiansen, P. Strength indicator values of theropod long bones, with comments on limb proportions and cursorial potential. *Gaia* **15**, 241–255 (2000).
- Farlow, J. O., Gatesy, S. M., Holtz, T. R. Jr, Hutchinson, J. R. & Robinson, J. M. Theropod locomotion. *Am. Zool.* **40**, 640–663 (2000).
- Irby, G. V. Paleontological evidence for running dinosaurs worldwide. *Mus. N. Arizona Bull.* **60**, 109–112 (1999).
- Coombs, W. P. Jr. Sauropod habits and habitats. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **17**, 1–33 (1975).
- Roberts, T. J. Muscle force and stress during running in dogs and wild turkeys. *Bull. Mus. Comp. Zool.* **156**, 283–295 (2001).
- Biewener, A. A. Scaling body support in mammals: limb posture and muscle mechanics. *Science* **245**, 45–48 (1989).
- Roberts, T. R., Chen, M. S. & Taylor, C. R. Energetics of bipedal running. II. Limb design and running mechanics. *J. Exp. Biol.* **201**, 2753–2762 (1998).
- Hutchinson, J. R. *The Evolution of Hindlimb Anatomy and Function in Theropod Dinosaurs*. PhD thesis, Univ. California (2001).
- Carrano, M. T. & Hutchinson, J. R. The pelvic and hind limb musculature of *Tyrannosaurus rex* (Dinosauria: Theropoda). *J. Morphol.* (in the press).
- Alexander, R. McN., Jayes, A. S., Maloij, G. M. O. & Wathuta, E. M. Allometry of the leg muscles of mammals. *J. Zool.* **194**, 539–552 (1981).
- Bennett, M. B. Allometry of the leg muscles of birds. *J. Zool.* **238**, 435–443 (1996).
- Grand, T. I. Body weight: its relation to tissue composition, segment distribution, and motor function. II. Interspecific comparisons. *Am. J. Phys. Anthropol.* **47**, 211–240 (1977).
- Hallgrímsson, B. & Maiorana, V. Variability and size in mammals and birds. *Biol. J. Linn. Soc.* **70**, 571–595 (2000).
- Gatesy, S. M. Guineafowl hind limb function. I: Cineradiographic analysis and speed effects. *J. Morphol.* **240**, 115–125 (1999).
- Johnston, I. A. Sustained force development: specializations and variation among the vertebrates. *J. Exp. Biol.* **115**, 239–251 (1985).
- Blickhan, R. The spring-mass model for running and hopping. *J. Biomech.* **22**, 1217–1227 (1989).
- Alexander, R. McN. The maximum forces exerted by animals. *J. Exp. Biol.* **115**, 231–238 (1985).
- Cavagna, G. A., Heglund, N. C. & Taylor, C. R. Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure. *Am. J. Physiol.* **233**, 243–261 (1977).
- Alexander, R. McN. & Jayes, A. S. A dynamic similarity hypothesis for the gaits of quadrupedal mammals. *J. Zool.* **201**, 135–152 (1983).
- Alexander, R. McN., Maloij, G. M. O., Njau, R. & Jayes, A. S. Mechanics of running of the ostrich (*Struthio camelus*). *J. Zool.* **187**, 169–178 (1979).
- Mendez, J. & Keys, A. Density and composition of mammalian muscle. *Metabolism* **9**, 184–188 (1960).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

We thank F. C. Anderson, K. Angielczyk, S. Delp, R. J. Full, S. M. Gatesy, T. M. Keaveny, R. Kram, K. Padian and J. F. Parham as well as the Berkeley Biomechanics and Stanford Neuromuscular Biomechanics seminars for input. J.R.H. appreciates funding from the University of California Department of Integrative Biology and Museum of Paleontology, and the National Science Foundation under a grant awarded in 2001. M.G. had postdoctorate funding through a DARPA grant to the Polypedal Lab at UC Berkeley. This is University of California Museum of Paleontology publication number 1749.

Correspondence and requests for materials should be addressed to J.R.H. (e-mail: [jrhutch@stanford.edu](mailto:jrhutch@stanford.edu)).



# Adaptive protein evolution in *Drosophila*

Nick G. C. Smith\*† & Adam Eyre-Walker\*

\* Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

For over 30 years a central question in molecular evolution has been whether natural selection plays a substantial role in evolution at the DNA sequence level<sup>1,2</sup>. Evidence has accumulated over the last decade that adaptive evolution does occur at the protein level<sup>3,4</sup>, but it has remained unclear how prevalent adaptive evolution is. Here we present a simple method by which the number of adaptive substitutions can be estimated and apply it to data from *Drosophila simulans* and *D. yakuba*. We estimate that 45% of all amino-acid substitutions have been fixed by natural selection, and that on average one adaptive substitution occurs every 45 years in these species.

Mutations can spread through a population either by random genetic drift or by the action of natural selection. The relative contributions of these two processes to evolution at the DNA level is one of the oldest and most keenly debated questions in molecular evolution<sup>1,2</sup>. Yet despite extensive analysis<sup>3</sup>, no consensus has been reached. With the great increase in single nucleotide polymorphism data, however, we are now in a position to tackle this question.

The proportion of adaptive mutations can be estimated by a simple extension of the McDonald–Kreitman test<sup>5–7</sup>. Let us begin by assuming that all synonymous mutations are neutral, and that all non-synonymous mutations are either strongly deleterious, neutral or strongly advantageous. Under this model the numbers of synonymous ( $P_s$ ) and non-synonymous ( $P_n$ ) polymorphisms segregating in a sample of sequences from a population are equal to  $4N_e u L_s k$  and  $4N_e u f L_n k$  respectively for an autosomal locus, where  $N_e$  is the effective population size<sup>2</sup>,  $u$  is the nucleotide mutation rate,  $f$  is the proportion of amino-acid mutations which are neutral,  $L_s$  and  $L_n$  are the numbers of synonymous and non-synonymous sites respectively and  $k$  is a constant reflecting the probability of observing a neutral variant. The constant  $k$  is dependent upon a number of factors including the number of sequences sampled, the sampling strategy, demography, background selection<sup>8</sup> and genetic hitch-hiking<sup>9</sup>; however, because synonymous and non-synonymous sites are interspersed, the value of  $k$  is the same for neutral mutations at each type of site. We assume that advantageous mutations contribute little to polymorphism, although they may contribute substantially to the divergence between species. This is not an unrealistic assumption; at most an advantageous mutation will contribute twice as much heterozygosity during its lifetime as a neutral variant<sup>2</sup>. For example, if advantageous mutations, with an advantage of  $N_e s = 25$  (where  $s$  is the strength of selection) occur at one-hundredth the rate of neutral mutations, they will account for 50% of substitutions, but account for just 2% of the heterozygosity. The numbers of synonymous ( $D_s$ ) and non-synonymous ( $D_n$ ) substitutions are  $2utL_s$ , and  $2utfL_n + a$ , where  $t$  is the time of divergence between the two species being considered (strictly the average time to coalescence of the genealogies of the sites being considered), and  $a$  is the number of adaptive substitutions. It is not difficult to show from these equations that the number of adaptive substitutions in a gene can be estimated by

$$a = D_n - D_s \frac{P_n}{P_s} \quad (1)$$

**Table 1** Proportion of amino-acid substitutions driven by positive selection

| Data set                                                                                                                                                                                   | x  | Number of genes | $\bar{\alpha}$ (95% C.I.) | Proportion of $\bar{\alpha} < 0$ |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----------------|---------------------------|----------------------------------|
| All genes                                                                                                                                                                                  | 0  | 35              | 0.24 (–0.11, 0.49)        | 0.083                            |
|                                                                                                                                                                                            | 5  | 30              | 0.43 (0.21, 0.61)         | 0.001                            |
|                                                                                                                                                                                            | 10 | 25              | 0.48 (0.30, 0.65)         | 0                                |
|                                                                                                                                                                                            | 20 | 12              | 0.52 (0.26, 0.74)         | 0                                |
|                                                                                                                                                                                            | 5  | 27              | 0.35 (0.08, 0.53)         | 0.011                            |
| Excluding <i>mth</i> , <i>Zw</i> and <i>Hex-t1</i>                                                                                                                                         |    |                 |                           |                                  |
| Genes with $P_s \leq x$ were excluded. The last column gives the proportion of bootstrap replicates in which $\bar{\alpha}$ was estimated to be less than zero. C.I., confidence interval. |    |                 |                           |                                  |

So dividing this expression by  $D_n$  gives an estimate of the proportion of amino-acid substitutions driven by positive selection

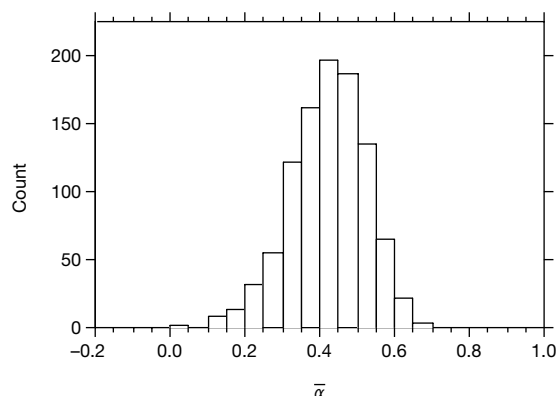
$$\alpha = 1 - \frac{D_s P_n}{D_n P_s} \quad (2)$$

To estimate the average proportion of amino-acid substitutions which are driven by adaptive evolution we need to combine data across genes. Unfortunately, both equations (1) and (2) are undefined if  $P_s = 0$ , and equation (2) is undefined if  $D_n = 0$ . Furthermore, caution must be exercised in summing the values of  $D_n$ ,  $D_s$ ,  $P_n$  and  $P_s$  across genes as this will give an overestimate of adaptive substitution if  $N_e$  and  $f$  are negatively correlated, as we might expect them to be:  $N_e$  is thought to vary across the genome in *Drosophila*<sup>10</sup> owing to processes such as genetic hitch-hiking<sup>9</sup> and background selection<sup>8</sup>, and this variation, in association with slightly deleterious mutations, will generate a negative correlation between  $N_e$  and  $f$ . We have therefore estimated the average proportion of amino-acid substitutions driven by positive selection by the expression

$$\bar{\alpha} = 1 - \frac{\bar{D}_s}{\bar{D}_n} \left( \frac{\bar{P}_n}{\bar{P}_s + 1} \right) \quad (3)$$

where all averages are across genes. We use the average of  $P_n/(P_s + 1)$  to estimate the mean value of  $L_n f/L_s$ , rather than the average of  $P_n/P_s$ , because  $P_n/(P_s + 1)$  is defined for all genes and is less biased; it is essentially unbiased if  $P_s > 5$ . The rationale behind equation (3) is given in the Methods section. The confidence interval of  $\bar{\alpha}$  was obtained by bootstrapping the data by randomly selecting genes with replacement.

We have used our method to estimate the proportion of amino-acid substitutions which have been driven by positive selection in the divergence between *D. simulans* and *D. yakuba* by using polymorphism data from *D. simulans* (see Supplementary Information for values of  $D_n$ ,  $D_s$ ,  $P_n$  and  $P_s$ ). To do this we compiled a data set of 43 genes for which we had multiple sequences from *D. simulans* and an orthologous *D. yakuba* sequence. From this data set we



**Figure 1** The distribution of 1,000 bootstrap values of  $\bar{\alpha}$  for the divergence between *Drosophila simulans* and *D. yakuba* for genes in which  $P_s > 5$ .  $\bar{\alpha}$  is the average proportion of amino-acid substitutions driven by positive selection.

† Present address: Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.

**Table 2 Numbers of synonymous and non-synonymous substitutions**

| Lineage                | $\bar{D}_n$ | $\bar{D}_s$ | $\bar{D}_n/\bar{D}_s$ |
|------------------------|-------------|-------------|-----------------------|
| <i>D. simulans</i>     | 6.15        | 14.40       | 0.43                  |
| <i>D. yakuba</i>       | 23.97       | 56.71       | 0.42                  |
| <i>D. melanogaster</i> | 10.91       | 18.26       | 0.60                  |

Substitutions are shown along the lineages leading to *D. simulans*, *D. yakuba* and *D. melanogaster* from the node which connects them. The ratio  $\bar{D}_n/\bar{D}_s$  is not significantly different between lineages.  $\bar{D}_n$  and  $\bar{D}_s$  are the average numbers of non-synonymous and synonymous substitutions.

discarded four genes that had been sequenced because they were thought to be likely targets of adaptive evolution, and four other genes because they contained no polymorphism in the *D. simulans* alleles. Using the remaining 35 genes we estimate that 24% of all amino-acid substitutions between *D. simulans* and *D. yakuba* have been driven by positive adaptive evolution (Table 1). This estimate is not significantly different from zero. However, some genes have very little polymorphism and therefore contribute a substantial amount of variance to the estimate of  $\bar{\alpha}$ . If we remove genes which have five or fewer synonymous polymorphisms the estimate of  $\bar{\alpha}$  is increased to 43%, which is significantly greater than zero (Table 1, Fig. 1). Removal of genes with more synonymous polymorphisms increases the estimate of  $\bar{\alpha}$  slightly to about 50%. We therefore estimate that  $\bar{\alpha}$  is about 45%. This is similar to an estimate recently obtained in primates:  $\bar{\alpha} = 35\%$  (ref. 6). However, the removal of genes with low  $P_s$  values has the potential to bias the estimate of  $\bar{\alpha}$  upwards; we therefore conducted a series of simulations which suggest that removing genes with low  $P_s$  values is not a problem in this data set (see Methods).

We noticed that our estimate of  $\bar{\alpha}$  is reduced only slightly if we remove three genes (*mth*, *Zw* and *Hex-1*) that individually show evidence of adaptive substitution when we conduct a McDonald–Kreitman test<sup>7</sup> (at  $P = 0.05$ , not correcting for multiple tests) (Table 1). Thus our method reveals evidence of adaptive evolution even in genes which do not individually show any evidence of positive selection.

However, the method makes two critical assumptions: (1) that the average proportion of amino-acid mutations which are neutral,  $\bar{f}$ , is constant through time; and (2) that the mutations segregating within a species are neutral. To test whether a change in  $\bar{f}$  could be responsible for the high estimate of  $\bar{\alpha}$  we used *D. melanogaster* to partition the values of  $\bar{D}_n$  and  $\bar{D}_s$  between the lineages leading to *D. simulans* and *D. yakuba* (see Methods). If there has been little or no adaptive evolution (that is,  $\bar{\alpha} = 0$ ) but  $\bar{f}$  was different in the past, then we expect the ratio  $\bar{D}_n/\bar{D}_s$  to differ between the lineages leading to *D. simulans* and *D. yakuba*. The ratio, however, is almost identical in those lineages, whereas it is substantially higher along the lineage leading to *D. melanogaster* (Table 2).

The method also makes the assumption that the mutations segregating as polymorphisms within the sample of sequences are neutral, whereas there is evidence that selection is acting upon synonymous mutations segregating in *D. simulans*<sup>11,12</sup>. However, our estimate of  $\bar{\alpha}$  will only be an overestimate if the synonymous mutations which are segregating are more deleterious on average than the non-synonymous mutations which are segregating. To investigate this, we calculated the difference between the average frequency of synonymous and the average frequency of non-synonymous polymorphisms segregating in each gene, taking the frequency of each polymorphism as the frequency of the minor allele. There is no apparent difference between the frequencies of two types of polymorphism in our sample (mean difference in frequencies is  $-0.004$ ,  $P > 0.1$  in a paired *t*-test), which suggests that synonymous and non-synonymous polymorphisms are subject to similar levels of selection, and that our estimate of  $\bar{\alpha}$  is unbiased.

We have estimated that approximately 45% of all amino-acid substitutions between *D. simulans* and *D. yakuba* have been adaptive. There are about 13,600 genes in the *Drosophila* genome, of average length of 590 codons<sup>13</sup>, and the average number of amino-

acid substitutions separating *D. simulans* and *D. yakuba* is 0.074 per codon (calculated from our data); we therefore estimate that there have been approximately 270,000 positively selected amino-acid substitutions in the evolution of *D. simulans* and *D. yakuba*. Given that *D. simulans* and *D. melanogaster* are thought, on the basis of biogeographic data, to have diverged 2.5 Myr ago<sup>14</sup>, we estimate from the data in Table 2 that *D. simulans* and *D. yakuba* diverged ~6 Myr ago. This implies that these two species have undergone one adaptive substitution every 45 years, or one substitution every 450 generations if *Drosophila* undergoes ten generations a year. This is consistent with Haldane's cost of natural selection<sup>15</sup>, the problem which first motivated Kimura to propose the neutral theory of molecular evolution<sup>16</sup>. □

## Methods

### Data

Collections of sequence data were assembled by searching the literature and sequence databases. If DNA sequence alignments were unavailable, we obtained our DNA alignments on the basis of protein alignments generated using the default parameters of ClustalW<sup>17</sup> as implemented in DAMBE version 4.0.31 (ref. 18). From our initial data set of 43 genes we excluded four genes because they segregated no polymorphisms in the sequences surveyed, and four genes that had been sequenced because they were thought to have undergone adaptive substitution; these latter genes were three reproductive genes (*janA*, *janB* and *ocr*) and one immune system gene (*relish*). We did not exclude genes that were sequenced because they segregated balanced polymorphisms or genes sequenced because they had high rates of non-synonymous substitution.

### Analysis

Polymorphisms were counted using DnaSP<sup>19</sup>. To determine the number of substitutions separating *D. simulans*, *D. yakuba* and *D. melanogaster* we aligned a single sequence from each species; if multiple alleles were available we randomly selected one. The number of synonymous and non-synonymous differences were calculated using the program *codeml*, as implemented in the PAML package<sup>20</sup> with codon frequencies estimated from the nucleotide frequencies at the three codon positions and the ratio of  $K_a/K_s$  free to vary down each lineage. (See the Supplementary Information for a list of genes along with their  $\bar{D}_n$ ,  $\bar{D}_s$ ,  $P_n$  and  $P_s$  values.)

To maintain consistency with the other analyses, McDonald–Kreitman tests were performed using the numbers of synonymous and non-synonymous differences calculated using PAML. The significance of the McDonald–Kreitman test was assessed by a Fisher's exact test; all genes which were significant at 5% showed an excess of non-synonymous substitution, rather than non-synonymous polymorphism.

### Rationale for equation (3)

Equation (3) can be expanded to a form which reflects the rationale better:

$\bar{\alpha} = (\bar{D}_n - \bar{D}_s(P_n/(P_s + 1)))/\bar{D}_n$ . The numerator estimates the average number of amino-acid substitutions driven by positive selection per gene, and the expression  $(\bar{P}_n/(P_s + 1))$  estimates the mean of  $L_{nf}/L_n$ . We use  $(\bar{P}_n/(P_s + 1))$  rather than  $(\bar{P}_n/\bar{P}_s)$  because the former is less biased—it is essentially unbiased if  $P_s > 5$ —and is defined for all values of  $P_s$ . The numerator is expected to be unbiased since  $E(\bar{D}_n - \bar{D}_s z) = E(\bar{D}_n) - E(\bar{D}_s)E(z)$  if  $\bar{D}_s$  and  $z = L_{nf}/L_n$  are uncorrelated ( $E(y)$  is the expected value of  $y$ ); we do not expect  $\bar{D}_s$  and  $z$  to be strongly correlated and there is no correlation in our data sets between  $\bar{D}_s$  and  $P_n/(P_s + 1)$  either for all genes ( $r^2 = 0.006$ ,  $P > 0.1$ ) or for genes for which  $P_s > 5$  ( $r^2 = 0.05$ ,  $P > 0.1$ ).

### Simulations

To investigate the bias associated with estimating  $\bar{\alpha}$  we conducted a simulation in which we used the observed values of  $P_s$  and  $\bar{D}_s$  to generate new data sets. Values of  $\bar{D}_n$ ,  $\bar{D}_s$ ,  $P_n$  and  $P_s$  were randomly generated from Poisson distributions with means of  $\bar{D}_s$ ,  $\bar{D}_s$ ,  $\bar{P}_s$  and  $\bar{P}_s$  respectively where  $\bar{D}_s$  and  $\bar{P}_s$  are the observed values of  $\bar{D}_s$  and  $P_s$  and  $z = L_{nf}/L_n$ . The value of  $\bar{\alpha}$  was calculated for each replicate data set, removing genes with  $P_s < x$  after randomly generating a data set to simulate the effect of removing genes with low  $P_s$ . We investigated three models: (1) a single value of  $z$  was applied to all genes, (2)  $z$  was allowed to vary randomly, and (3)  $z$  was negatively correlated to  $P_s$ . The value of  $\alpha$  was set to zero in each gene. For each set of parameters (that is, values of  $z$  and the value of  $P_s$  below which genes are discarded) 100 replicate data sets were produced. We calculated the mean of  $\bar{\alpha}$  and the 95% percentiles. In simulations with  $P_s > 5$  the mean of  $\bar{\alpha}$  was less than 5% in all simulations with the 95% percentile being below 13% in all simulations (see the Supplementary Information for additional simulation results).

Received 15 May; accepted 19 December 2001.

- Gillespie, J. H. *The Causes of Molecular Evolution* (Oxford Univ. Press, Oxford, 1991).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
- Kreitman, M. & Akashi, H. Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* **26**, 403–422 (1995).
- Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).

5. Charlesworth, B. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**, 213–227 (1994).
6. Fay, J., Wyckoff, G. J. & Wu, C.-I. Positive and negative selection on the human genome. *Genetics* **158**, 1227–1234 (2001).
7. McDonald, J. H. & Kreitman, M. Adaptive evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
8. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
9. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
10. Begun, D. J. & Aquadro, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520 (1992).
11. Begun, D. The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18**, 1343–1352 (2001).
12. Kliman, R. Recent selection on synonymous codon usage in *Drosophila*. *J. Mol. Evol.* **49**, 343–351 (1999).
13. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
14. Powell, J. R. & DeSalle, R. *Drosophila* molecular phylogenies and their uses. *Evol. Biol.* **28**, 87–138 (1995).
15. Haldane, J. B. S. The cost of natural selection. *J. Genet.* **55**, 511–524 (1957).
16. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
17. Thompson, J. D., Higgins, D. G. & Gibson, T. J. ClustalW—improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680 (1994).
18. Xia, X. *Data Analysis in Molecular Biology and Evolution* (Kluwer Academic, London, 2000).
19. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).
20. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

We thank B. Charlesworth, C.-I. Wu, S. Otto, M. Whitlock, T. Johnson, P. Awadalla, J. Gillespie, G. McVean and P. Keightley for helpful discussions, and E. Moriyama for help with data collection. N.G.C.S. was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and A.E.-W. is funded by the Royal Society and the BBSRC.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E.-W. (e-mail: [a.c.eyre-walker@sussex.ac.uk](mailto:a.c.eyre-walker@sussex.ac.uk)).

# Testing the neutral theory of molecular evolution with genomic data from *Drosophila*

Justin C. Fay<sup>\*†</sup>, Gerald J. Wyckoff<sup>\*†</sup> & Chung-I Wu<sup>\*‡</sup>

<sup>\*</sup> Committee on Genetics, University of Chicago, Chicago, Illinois 60637, USA

<sup>‡</sup> Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA

Although positive selection has been detected in many genes, its overall contribution to protein evolution is debatable<sup>1</sup>. If the bulk of molecular evolution is neutral, then the ratio of amino-acid (A) to synonymous (S) polymorphism should, on average, equal that of divergence<sup>2</sup>. A comparison of the A/S ratio of polymorphism in *Drosophila melanogaster* with that of divergence from *Drosophila simulans* shows that the A/S ratio of divergence is twice as high—a difference that is often attributed to positive selection. But an increase in selective constraint owing to an increase in effective population size could also explain this observation, and, if so, all genes should be affected similarly. Here we show that the difference between polymorphism and divergence is limited to only a

fraction of the genes, which are also evolving more rapidly, and this implies that positive selection is responsible. A higher A/S ratio of divergence than of polymorphism is also observed in other species, which suggests a rate of adaptive evolution that is far higher than permitted by the neutral theory of molecular evolution.

The neutral theory holds that the bulk of DNA divergence between species is driven by mutation and drift, rather than by positive darwinian selection<sup>3</sup>. But because the effect of positive selection is often masked by negative selection<sup>4</sup>, detecting positive selection is a challenging task. A rate of amino-acid substitution greater than that of synonymous substitution can be explained only by positive selection<sup>5</sup>, but such a criterion is very stringent as negative selection lowers the rate of amino-acid substitution. A high rate of amino-acid substitution is limited mostly to genes that are involved in resistance to disease or in sexual reproduction, where there is continual room for improvement<sup>6,7</sup>.

The McDonald–Kreitman test can detect positive selection even in the presence of negative selection through a ratio of amino-acid divergence to synonymous divergence greater than that of polymorphism<sup>2</sup>. The A/S ratio of divergence is inflated above polymorphism by advantageous amino-acid mutations, which quickly sweep through a population but have a cumulative effect on divergence. The McDonald–Kreitman test has been applied to many genes individually, but only a few have yielded a significant excess of amino-acid divergence (*Drosophila* genes are reviewed in refs 8, 9). This may in part be caused by a lack of power in detecting positive selection in individual genes unless a large number of adaptive substitutions have occurred.

For those genes that have yielded a significant McDonald–Kreitman test result, the A/S ratio of divergence is more than twice as great as polymorphism<sup>10–12</sup>. The effects of positive selection may also be obscured by slightly deleterious amino-acid mutations that inflate the A/S ratio of polymorphism but not divergence. The effects of slightly deleterious mutations can be removed by comparing common polymorphism with divergence, because deleterious amino-acid mutations are kept at low frequency in the population<sup>4</sup>. This can only be done when the data from a large number of genes are combined; individual genes rarely contain more than a few common amino-acid polymorphisms.

An important but rarely appreciated assumption of the McDonald–Kreitman test is that the selective constraint on a gene remains constant over time. The selective constraint on a gene is determined by the proportion of amino-acid mutations that are deleterious<sup>3</sup>,  $2N_s < -1$ , so both a change in the selection coefficient ( $s$ ) and a change in effective population size ( $N$ ) can result in a change in selective constraint. Although it is well known that selective constraint is not static across phylogenetic lineages<sup>13,14</sup>, this assumption is rarely justified in applications of the McDonald–Kreitman test. Whereas the strength of selection on each gene might fluctuate over time depending on the genetic or environmental background, a genome-wide change in constraint, such as that caused by a change in effective population size, should produce a consistent increase or decrease in the A/S ratio across all genes. Alternatively, under positive selection each gene might be affected to a different degree and some genes might not be affected at all.

To compare genomic patterns of amino-acid and synonymous

**Table 1 Polymorphisms in *D. melanogaster* and divergence from *D. simulans***

| Gene*     | Class                  | Amino-acid polymorphism, A | Synonymous polymorphism, S | A/S  |
|-----------|------------------------|----------------------------|----------------------------|------|
| X-linked  | Rare ( $\leq 12.5\%$ ) | 4                          | 67                         | 0.06 |
|           | Common ( $> 12.5\%$ )  | 6                          | 46                         | 0.13 |
|           | Divergence             | 42                         | 189                        | 0.22 |
| Autosomal | Rare                   | 79                         | 126                        | 0.63 |
|           | Common                 | 44                         | 118                        | 0.37 |
|           | Divergence             | 421                        | 521                        | 0.81 |

\* There are 5 X-linked and 31 autosomal genes with a sample size of eight or greater (see text for the data from all 45 genes).

<sup>†</sup> Present addresses: Department of Genome Sciences, Lawrence Berkeley National Laboratory, Berkeley, California 94720 (J.C.F.); Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA (G.J.W.).



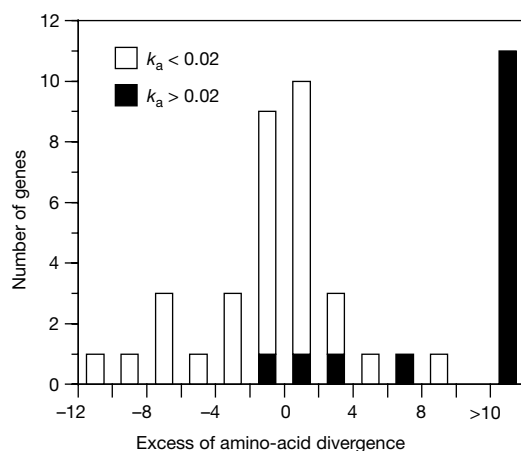
**Table 2 African and non-African common polymorphism and divergence**

| Class        | Population  | Amino-acid polymorphism, A | Synonymous polymorphism, S | A/S  |
|--------------|-------------|----------------------------|----------------------------|------|
| Polymorphism | Non-African | 48                         | 124                        | 0.39 |
|              | African     | 40                         | 159                        | 0.25 |
| Divergence   |             | 413                        | 663                        | 0.62 |

site evolution, we tabulated polymorphism in *D. melanogaster* and divergence from *D. simulans* from 45 gene surveys (Methods). If all amino-acid and synonymous variation is neutral, then the A/S ratio of polymorphism and divergence should be constant. The A/S ratio of divergence ( $598/950 = 0.63$ ) is significantly greater than that of common polymorphism ( $65/224 = 0.29$ ;  $P < 10^{-6}$ ). We compared divergence with the common rather than the total polymorphism because deleterious mutations at low frequency inflate the A/S ratio of polymorphism. For the 36 genes with sample sizes of eight or greater, there is a significant excess of rare over common amino-acid variation in autosomal genes ( $P = 0.022$ ; Table 1), as is observed in humans<sup>4</sup>. The absence of a difference in X-linked genes suggests that the deleterious mutations are partially recessive and are more readily eliminated from the X chromosome.

Both positive selection and an increase in selective constraint on amino-acid changes can produce a higher A/S ratio of divergence than of polymorphism. But only under certain restrictive conditions is a genome-wide change in constraint possible. One such condition is an increase in effective population size that is neither too distant nor too recent in the evolutionary past. If this possibility can be ruled out, positive selection may be the only viable explanation for the high rate of amino-acid divergence.

If an increase in selective constraint resulted from a population size increase associated with the spread of *D. melanogaster* outside Africa<sup>15</sup>, it might be more appropriate to compare the A/S ratio of the African population with that of divergence. Table 2, which includes the 32 genes for which both African and non-African populations were surveyed, shows that there is a significantly larger A/S ratio of divergence than of polymorphism in either population. If a recent increase in effective population size increased constraint on amino-acid polymorphism in both African and non-African populations, then patterns of synonymous polymorphism might be skewed towards rare variants. Neither African or non-African populations show this pattern<sup>16</sup>. Finally, if there has been a decrease in effective population size along the *D. melanogaster* lineage<sup>17,18</sup>, the A/S ratio of polymorphism should be greater than that of divergence between the two species.



**Figure 1** The distribution of the excess of amino-acid divergence contributed by each gene. For reference, fast and slowly evolving genes are denoted by a rate of amino-acid substitution ( $k_a$ ) greater than (filled bars) or less than (open bars) 2%.

**Table 3 Polymorphism and divergence in neutral and fast genes**

| Genes*  | Class      | Amino-acid polymorphism, A | Synonymous polymorphism, S | A/S  |
|---------|------------|----------------------------|----------------------------|------|
| Neutral | Rare       | 31                         | 90                         | 0.34 |
|         | Common     | 16                         | 69                         | 0.23 |
|         | Divergence | 65                         | 247                        | 0.26 |
| Fast    | Rare       | 48                         | 36                         | 1.33 |
|         | Common     | 28                         | 49                         | 0.57 |
|         | Divergence | 356                        | 274                        | 1.30 |

\*X-linked genes are excluded.

If an increase in effective population size has produced a genome-wide increase in selective constraint, the A/S ratio of all genes should be affected. In Fig. 1, the distribution of each gene's contribution to the excess of amino-acid divergence suggests that there are two classes of gene: neutral and rapidly evolving. The neutral class comprises 34 genes that deviate by less than 10 amino-acid substitutions from that expected on the basis of the A/S ratio of all common polymorphism. The remaining 11 genes all have a higher A/S ratio of divergence than of polymorphism, and account for the whole difference in the A/S ratio of polymorphism and divergence. These genes are *Acp26Aa*, *Acp29Ab*, *anon1A3*, *anon1E9*, *anon1G5*, *ci*, *est-6*, *Ref2P*, *Rel*, *tra* and *Zw*. As expected under positive selection, which increases the rate of protein evolution, these 11 genes have a high rate of amino-acid substitution (Fig. 1).

Can the pattern in Fig. 1 be explained by selection or demography? Table 3 shows that, in the rapidly evolving genes, the A/S ratios of divergence and of rare polymorphism are much higher than the A/S ratio of the common polymorphism. This is expected if the genes are under positive selection. Although a large increase in population size in the recent past could account for the difference between the A/S ratio of divergence and that of common polymorphism, this explanation is incompatible with the very small difference found in the 26 neutral genes. Because both the neutral and rapidly evolving genes have a higher A/S ratio of rare polymorphism than of common polymorphism, both should have been affected by an increase in effective population size.

If positive selection is common, other species should also have an A/S ratio of divergence greater than that of polymorphism. In addition, any demographic scheme is not likely to be shared by several species. In a study of eight genes in *D. simulans*, *Drosophila mauritiana* and *Drosophila sechellia*, the A/S ratio of polymorphism ( $A/S = 32/183$ ) is 34% that of divergence ( $28/55$ )<sup>19</sup>. In a study of 42 genes with polymorphism in both *D. melanogaster* and *D. simulans*, the A/S ratio of polymorphism is 65% that of divergence (N. G. C. Smith and A. Eyre-Walker, personal communication). In another study of 23 genes, the A/S ratio of polymorphism ( $45/305$ ) is 30% that of divergence along the *D. simulans* lineage ( $65/133$ )<sup>20</sup>. In humans, the A/S ratio of common polymorphism ( $70/122$ ) found in 181 genes is 65% that of divergence ( $3,660/4,151$ ) found in a different set of 182 human and Old World monkey genes<sup>4</sup>.

Although these genomic patterns of variation are not explained easily by the neutral theory, slightly deleterious mutations must clearly be accounted for in attempting to measure positive selection. In humans, 38% of amino-acid polymorphism was estimated to be slightly deleterious<sup>4</sup>, and in *D. melanogaster* the estimate is 26%,  $(0.63 - 0.37) \times 126/123$ , from the combined neutral and rapidly evolving genes (Table 3). These slightly deleterious mutations, which are emphasized by the nearly neutral theory<sup>21</sup>, could become effectively neutral and fixed during a population bottleneck of sufficient severity, providing a burst of amino-acid substitutions and an increase in the A/S ratio of divergence. We control for the impact of these slightly deleterious mutations by comparing the rapidly evolving class of gene to the neutral class (Fig. 1, Table 3). Additional genomic data from other species will be needed to estimate the general impact of these slightly deleterious mutations on protein evolution. □

## Methods

### Data

A literature search yielded 45 genes for which polymorphism had been surveyed in *D. melanogaster* and for which an outgroup sequence was available. Of these, 36 had a sample size of eight or greater, 32 had been surveyed in at least two African and two non-African individuals and 10 were of X-linked genes. The 45 genes and their references are listed in Supplementary Information.

### Analysis

Polymorphism data was tabulated by hand or from GenBank accession numbers using SITES<sup>31</sup> or DNASP<sup>32</sup>. For each polymorphic site, the minor allele was classified as rare ( $\leq 12.5\%$ ) or common ( $> 12.5\%$ ). The cutoff of 12.5% was chosen to exclude deleterious mutations from the common frequency class and to include those genes with samples of eight or more in the analysis of rare compared to common polymorphism. Cutoffs of 10 and 15% produce similar results. We treated three alleles segregating at a single nucleotide as two segregating sites and excluded complex variations. Divergence data was obtained by comparing a randomly chosen sequence of *D. melanogaster* with that of *D. simulans* or, if unavailable, either *D. mauritiana* or *D. sechellia*. The number of amino-acid and synonymous substitutions between species was estimated using Kimura's two-parameter model to correct for multiple hits.

The contribution of each gene to the excess number of amino-acid substitutions was calculated as the excess number of amino-acid substitutions minus the excess number of amino-acid polymorphisms found in each gene. The excess for polymorphism and divergence is  $A - S \times (65/224)$ , where  $A$  and  $S$  are the number of amino-acid and synonymous substitutions, respectively, and 65/224 is the total number of amino-acid polymorphisms divided by synonymous polymorphisms. (Ideally, the excess of amino-acid divergence in each gene should be calculated using only polymorphism and divergence in that gene but there is rarely sufficient polymorphism in a single gene for comparison with divergence.) We also calculated the contribution to the excess separately for three groups of genes sorted by their rate of amino-acid divergence. The two methods produced a similar distribution so the simpler method using a single group of genes was used.

Received 27 June; accepted 4 December 2001.

1. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
2. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
3. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
4. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Positive and negative selection on the human genome. *Genetics* **158**, 1227–1234 (2001).
5. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).
6. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
7. Wyckoff, G. J., Wang, W. & Wu, C.-I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309 (2000).
8. Weinreich, D. M. & Rand, D. M. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* **156**, 385–399 (2000).
9. Moriyama, E. N. & Powell, J. R. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**, 261–277 (1996).
10. Eanes, W. F., Kirchner, M. & Yoon, J. Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. *Proc. Natl Acad. Sci. USA* **90**, 7475–7479 (1993).
11. Begun, D. J. & Whitley, P. Adaptive evolution of *relish*, a *Drosophila* NF- $\kappa$ B/I $\kappa$ B protein. *Genetics* **154**, 1231–1238 (2000).
12. Tsur, S. C., Ting, C. T. & Wu, C. I. Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*. II. Divergence versus polymorphism. *Mol. Biol. Evol.* **15**, 1040–1046 (1998).
13. Langley, C. H. & Fitch, W. M. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177 (1974).
14. Ohta, T. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**, 56–63 (1995).
15. Lachaise, D. M., Cariou, M.-L., David, J. R., Lemeunier, F. & Tsacas, L. The origin and dispersal of the *Drosophila melanogaster* subgroup: a speculative paleogeographic essay. *Evol. Biol.* **22**, 159–225 (1988).
16. Andolfatto, P. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**, 279–290 (2001).
17. Akashi, H. Codon bias evolution in *Drosophila*: Population genetics of mutation-selection drift. *Gene* **205**, 269–278 (1997).
18. McVean, G. A., Vieira, J. Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* **157**, 245–257 (2001).
19. Kliman, R. M. *et al.* The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931 (2000).
20. Begun, D. J. The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18**, 1343–1352 (2001).
21. Ohta, T. Slightly deleterious mutant substitutions during evolution. *Nature* **246**, 96–98 (1973).
22. Hey, J. & Wakeley, J. A coalescent estimator of the population recombination rate. *Genetics* **145**, 833–846 (1997).
23. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

This work was supported by grants from the NIH and NSF to C.-I.W. and a Genetics Training Grant and a Department of Education PhD fellowship to J.C.F.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C.F. (e-mail: jcfay@lbl.gov).

# Brain potential and functional MRI evidence for how to handle two languages with one brain

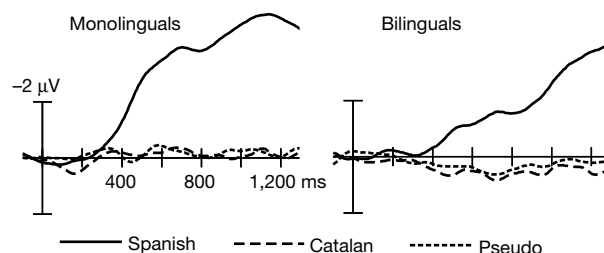
Antoni Rodríguez-Fornells\*, Michael Rotte†, Hans-Jochen Heinze†, Tömmé Nösselt† & Thomas F. Münte\*

\* Department of Neuropsychology, Otto von Guericke University, Universitätsplatz 2, Gebäude 24, 39106 Magdeburg, Germany

† Klinik für Neurologie 2, Otto von Guericke University, Leipzigerstrasse 44, 39120 Magdeburg, Germany

Bilingual individuals need effective mechanisms to prevent interference from one language while processing material in the other<sup>1</sup>. Here we show, using event-related brain potentials and functional magnetic resonance imaging (fMRI), that words from the non-target language are rejected at an early stage before semantic analysis in bilinguals. Bilingual Spanish/Catalan and monolingual Spanish subjects were instructed to press a button when presented with words in one language, while ignoring words in the other language and pseudowords. The brain potentials of bilingual subjects in response to words of the non-target language were not sensitive to word frequency, indicating that the meaning of non-target words was not accessed in bilinguals. The fMRI activation patterns of bilinguals included a number of areas previously implicated in phonological and pseudoword processing<sup>2–5</sup>, suggesting that bilinguals use an indirect phonological access route to the lexicon of the target language to avoid interference<sup>6</sup>.

High-proficiency bilingual subjects manage to understand and speak one of their languages without apparent interference from the other. This is a remarkable ability in the face of the fact that neuroimaging studies have revealed, at least for high-proficiency bilinguals, that neuro-anatomical representations of both languages are



**Figure 1** Lateralized readiness potentials (LRPs) from the main experiment indicating the preparation of motor responses. The onset latency of the LRP to Spanish words, estimated by the time at which the amplitude was significantly different from zero for at least 4 consecutive time points (sequential  $t$ -tests)<sup>14</sup>, was 408 ms in the monolingual and 520 ms in the bilingual group. No LRP activity is observed for Catalan words, indicating an effective blocking of 'word' (go) responses in the bilingual group.

overlapping<sup>7,8</sup>. How, then, do bilinguals avoid interference between languages and maintain their attention focussed on the target language? In an experiment involving Spanish/English bilinguals who were shown a mixed list of Spanish, English and pseudowords (subjects were instructed to respond to words in the target language and reject both non-target words and pseudowords), non-target words were rejected as fast as pseudowords<sup>9</sup>. Furthermore, in contrast to words in the target language, no effect of word frequency on the rejection latencies of words in the non-target language was found<sup>9</sup>. This suggests that bilinguals effectively 'filter out' words of the non-target language, and are able selectively to shut down one lexicon when necessary<sup>10,11</sup>. However, other behavioural data have been thought to be compatible with a non-selective access to the two lexicons in bilinguals<sup>12</sup>.

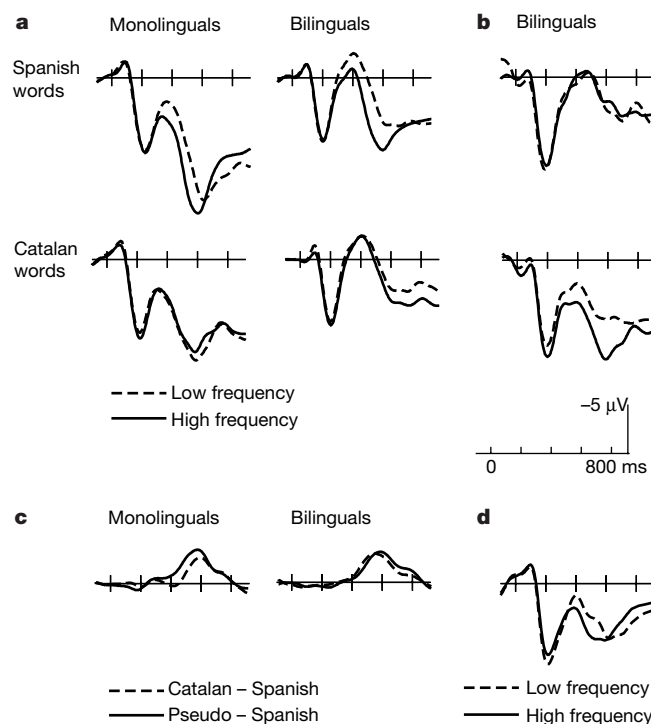
Thus, although it has been suggested that an automatic switching mechanism permits a rapid change between languages<sup>1</sup>, this notion is ill-defined, and it is unclear to what degree words from the non-target language are processed by bilinguals.

To address this question we assessed behavioural, electrophysiological and brain-imaging measures during word processing in a group of bilingual Catalan/Spanish-speaking subjects, and in a second group of monolingual speakers of Spanish. Catalan, as with Spanish, is a Romance language and is mainly spoken in the north-eastern part of Spain<sup>13</sup>. It is one of two official languages in this region and is compulsory in school and at university. In our experiments, subjects were required to make a speeded button press in response to Spanish words presented on a video monitor, with response hand (left or right) determined by the first letter (vowel or consonant). Subjects were instructed to withhold responses to Catalan words and pseudowords. Both groups performed the task

with ease. The difference in reaction times to Spanish words (monolingual, 699 ms; bilingual, 738 ms,  $F_{1,28} = 1.9$ ,  $P = 0.17$ ) failed to reach significance, but the delayed onset latency of the lateralized readiness potential (LRP)—which indexes response preparation<sup>14</sup>—in bilingual subjects (Fig. 1) suggests that bilinguals were slower in preparing responses to words in the target language. Bilinguals also showed slightly more false-positive responses to high-frequency Catalan words (monolingual: high frequency, 3.8%, low frequency, 3.9%; bilingual: high frequency, 5.7%, low frequency, 3.7%; group by frequency interaction  $F_{1,28} = 7.64$ ,  $P < 0.01$ ), indicating that they had some difficulty suppressing button presses to high-frequency irrelevant words.

For the target (Spanish) words, event-related brain potentials (ERPs) showed a typical sensitivity to word frequency in both groups, as shown in Fig. 2. This effect results from a modulation of the 'N400' component<sup>15</sup>, the frequency sensitivity of which is well known<sup>16</sup>—it receives contributions from semantic areas of the fusiform gyrus<sup>17</sup> and has been interpreted as reflecting meaning access<sup>18</sup>. Interestingly, neither of the groups showed frequency sensitivity for Catalan words. Also, ERP effects towards Catalan and pseudowords were virtually identical in both groups (Fig. 2c). This suggests that the meanings of most of the Catalan words are not accessed even in subjects who are highly proficient in that language, and that these words are rejected at a stage before semantic analysis.

To test the generality of these effects, a subset of the bilingual group was subjected to a control experiment with the instruction to press the button in response to Catalan words, and to ignore Spanish words and pseudowords. The pattern of ERPs was reversed, with a frequency effect now showing up for the Catalan words, whereas no frequency effect was seen for the Spanish stimuli



**Figure 2** Event-related brain potentials (ERPs) from the main and two control experiments. **a**, ERPs from the main experiment: only Spanish words show a frequency sensitivity with an increased negativity for low-frequency words. No frequency effect is seen for the Catalan words (mean amplitude 350–500 ms, language  $\times$  frequency interaction,  $F_{1,28} = 70.8$ ,  $P < 0.001$ , centroparietal electrode sites). **b**, ERPs from the first control experiment performed in a subset of the bilingual group. A frequency effect is only seen for the (relevant) Catalan words (task  $\times$  language  $\times$  frequency interaction,  $F_{1,3} = 37.05$ ,  $P < 0.01$ , task defines the relevant language). **c**, Difference waves obtained by subtracting the ERPs to the Spanish words from the ERPs to the Catalan words and

pseudowords, respectively. The resulting waveforms are very similar for both conditions, and for the Catalan group are virtually overlapping. Neither main effects of stimulus type (Catalan versus pseudowords) nor group by stimulus type interaction effects were obtained for ANOVA tests conducted on several successive 200-ms time windows between 100 and 700 ms. **d**, ERPs from the second control experiment. The frequency sensitivity of the ERPs in the 350–500-ms time window was preserved ( $F_{1,11} = 17.9$ ,  $P < 0.0014$ , centroparietal electrode sites), even though only pseudowords required a response in this experiment. All waveforms are from the central midline site. Scale bar applies to all panels.

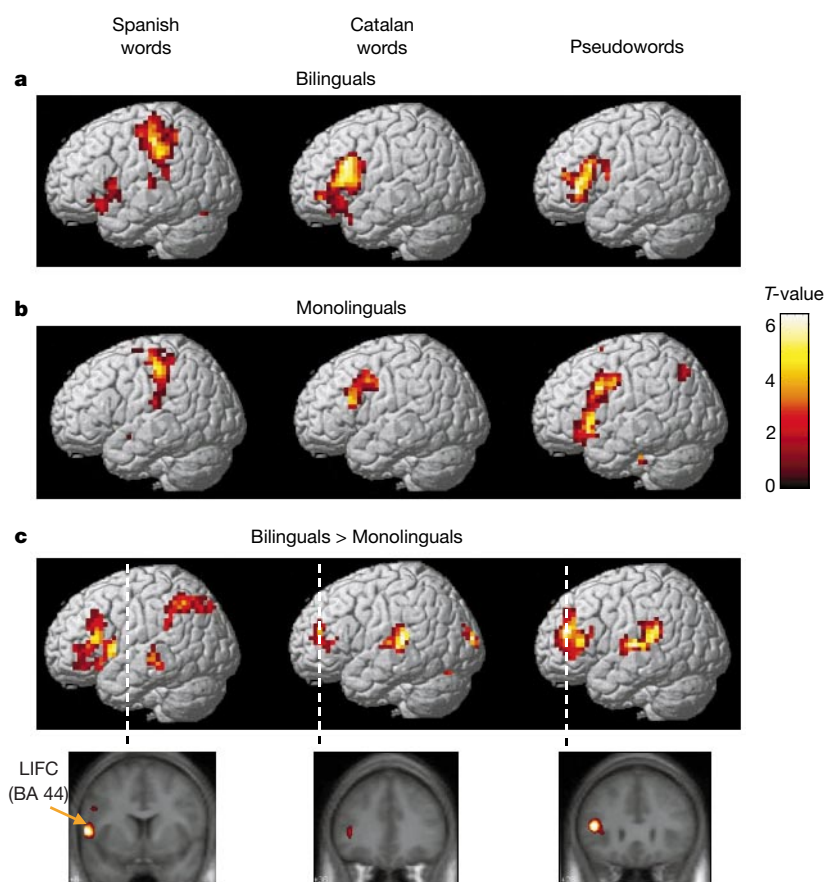


(Fig. 2b). To rule out the possibility that the lack of an ERP frequency effect for words from the non-target language could be due to the task characteristics, a second control experiment was conducted on monolingual Spanish subjects. Here, a discriminative (vowel/consonant) response was required for pseudowords occurring intermixed with Spanish words, which required no response (no-go). The frequency modulation of the N400 component was preserved for the Spanish words, indicating its independence of task requirements (Fig. 2d).

To examine the areas of the brain that are active during the task, we conducted an event-related functional MRI version of the experiment. Importantly, neither the monolingual nor the bilingual group showed reliable differences between pseudowords compared with Catalan words. The contrast between the three critical conditions (Spanish, Catalan and pseudowords) against a consonant string baseline (Fig. 3a, b) indicates how bilinguals may optimize performance during this task. Using responses to consonant strings (for example, dfmvr) as a baseline, activity in extrastriate visual cortex, which is usually found when contrasting words to a fixation baseline condition, is subtracted out<sup>19</sup>. Activity in the left primary motor cortex, left inferior parietal lobe and cerebellum, initiated by Spanish words, is attributable to the motor response. Only bilingual subjects exhibited an activation of the posterior inferior frontal area in response to the stimuli of Spanish words. This is evident when the pattern of activation between monolingual and bilingual groups is compared (Fig. 3c). This region of the brain is critical for pseudo-word reading<sup>20</sup>, phonological processing<sup>2,3</sup> and subvocal rehearsal<sup>4</sup>,

and was activated to a greater extent for Catalan (non-target) words and pseudowords in both groups. Furthermore, bilingual subjects showed a greater activation of the planum temporale (BA 22, 42) for all three critical conditions, a region of the brain that has also been linked to phonological processing<sup>2–5</sup>.

A parsimonious interpretation of these results, in line with the electrophysiological data, can be derived from the dual route model of reading, which proposes two different pathways for lexical access: the 'lexical' route, which proceeds directly from orthography to lexical access, and the 'sublexical' or letter-to-sound route, in which the graphemic form is converted to its phonological representation before lexical access, by applying a specific set of graphemic–phonological spelling rules<sup>5</sup>. We propose that bilinguals perform the current task using the sublexical pathway, that is, activating only graphemic–phonological spelling rules of the Spanish language, and that use of this pathway is reflected in a greater activation of posterior inferior frontal cortex and the planum temporale. This would imply an inhibition of the direct access route from orthography to the lexicon, thereby reducing response conflict for non-target words. The greater activation in an anterior prefrontal region (BA45/9, Fig. 3c) of the brain in bilingual subjects in response to Catalan words and pseudowords is a likely correlate of this inhibition, as this region has recently been implicated in the selection of relevant information and interference resolution<sup>21,22</sup>. These stimuli have also been associated with an enhanced activation in the anterior cingulate cortex that has frequently been found in control and conflict tasks<sup>23</sup> (see Supplementary Information). However,



**Figure 3** Rendering of group-averaged brain activations in standard stereotactic space identified for the Go condition (Spanish words) and the two no-go conditions (Catalan words and pseudowords). Each condition was contrasted against the baseline condition (consonant strings). **a**, The bilingual group ( $n = 7$ ); **b**, the monolingual group ( $n = 7$ ) ( $P < 0.01$ , 20 voxels). **c**, Group differences were identified using a  $t$ -test between groups for each of the conditions ( $P < 0.02$ , 40 voxels). Shown are regions with greater

activations in the bilinguals. Corresponding coronal views superimposed on the mean anatomical image of the 14 participants are shown (bottom). The group comparison for Spanish words showed a reliable difference in the left posterior inferior frontal cortex (LIFC, BA 44; peak coordinates  $-60, 8, 8$ ,  $P < 0.001$ , uncorrected). In the pseudoword condition the peak activity was located more anteriorly, near to the medial frontal gyrus (BA 45/9;  $-44, 28, 8$ ,  $P < 0.001$ ).

sometimes inhibition seems to have failed in our subjects, as demonstrated by the slight increase in commission errors for high-frequency words in the non-target language. Indeed, high-frequency words are thought to be processed using the direct access route<sup>6</sup>.

Taken together, behavioural, electrophysiological and brain-imaging measures demonstrate a very efficient blocking of the non-target language in bilinguals, which is probably achieved through the use of an indirect access route to the lexicon. The meaning of stimuli in the irrelevant language seems not to be accessed under the current conditions. Further studies are needed to test the generality of these findings in other experimental tasks. □

## Methods

### Subjects

Monolingual speakers of Spanish and bilingual Catalan/Spanish subjects (age 18–30 years) were recruited from the foreign student population at Hannover and Magdeburg Universities, Germany. All bilinguals had acquired both languages during their first years of life, and a questionnaire<sup>24</sup> revealed a high proficiency in and regular usage of both languages.

### ERP experiments

Fifteen monolingual and fifteen bilingual subjects participated in the main experiment. Six hundred Spanish words (300 high-frequency words, 95.0 occurrences per million; 300 low-frequency words, 2.45 per million<sup>25</sup>), 300 Catalan words (150 high-frequency words, 68.4 per million, 150 low-frequency words, 7.4 per million<sup>26</sup>) and 300 pseudowords (150 each derived from Spanish and Catalan by changing 1–3 letters) were presented in random order on a computer monitor (duration 400 ms, stimulus-onset asynchrony 1,750–2,250 ms). A fixation point appeared at the beginning of each trial in the centre of the screen 500 ms before imperative stimuli. Only Catalan words that were completely different (non-cognates) from Spanish words were selected. A speeded button press was required for Spanish words, whereas no response was to be given to Catalan and pseudowords. Spanish words with a vowel as the initial letter required a right-hand response, those starting with a consonant required a left-hand response (counterbalanced across subjects). This task was chosen to allow for the computation of the lateralized readiness potential (LRP). Several lists with different orders were generated and presented to subsets of the subjects. Event-related brain potentials (ERPs) and LRPs were computed from the electroencephalogram (EEG) using standard procedures<sup>14</sup>. Mean amplitude measures were obtained and entered into analysis of variance (ANOVA) statistics with the Huynh–Feldt epsilon correction applied as necessary. Before statistical analysis, error rates were replaced by the inverse sine transformation of the square root of the error proportions<sup>27</sup>.

In the first control experiment (four bilingual subjects), the Catalan words used were the same as in the main experiment. These were presented together with a subset of the Spanish words (75 high- and 75 low-frequency words; 90 and 2.4 per million, respectively) and 150 pseudowords. The instruction was the same as in the main experiment, except that Catalan words were now to be responded to with the first letter (vowel or consonant) determining response hand.

In the second control experiment (12 additional monolingual Spanish subjects), 300 Spanish words (150 of each frequency level) and 300 pseudowords (all derived from Spanish words) were used with a left/right reaction to pseudowords, depending on the initial letter (vowel or consonant).

### fMRI experiment

The fMRI session was conducted with the same timing as in the ERP session, but the number of stimuli was reduced (324 Spanish words, 108 Catalan words, 108 pseudowords and 108 consonant strings). The consonant strings were presented for further differentiation, and additional trials comprising the presentation of just a cross-hair served as a baseline condition. Imaging was performed with a GE Medical Systems 1.5 Tesla Signa Neurovascular MR scanner with a standard quadrature head coil. Visual images were back-projected onto a screen by an LED projector and participants (7 bilingual and 7 monolingual) viewed the images through a mirror on the head coil. Magnet-compatible response buttons were used. Conventional high-resolution structural images (rf-spoiled GRASS sequence, 60 sagittal slices, 2.8 mm thickness) were followed by functional images sensitive to blood oxygenation level-dependent contrast (echo planar  $T_2^*$ -weighted gradient echo sequence, repeat time = 2 s, echo time = 40 ms, flip = 90°). Each functional run consisted of 135 sequential whole-brain acquisitions (16 axial slices aligned to the plane intersecting the anterior and posterior commissures, 3.125-mm in-plane resolution, 7 mm thickness, 1 mm gap between slices). There were 4 trial types as in the ERP experiment plus the fixation control (cross-hair), and subjects performed 6 runs (mean stimulus-onset asynchrony of 2 s). Volumes were acquired continuously; the first four volumes were discarded to allow for  $T_1$  equilibration effects. To allow precise co-registration of functional data a separate  $T_1$ -weighted echo-planar image was acquired covering the whole volume (inversion recovery prepared echo-planar imaging (EPI) sequence). Different preprocessing steps were implemented using statistical parametric mapping (SPM99)<sup>28</sup>. Functional volumes were realigned, corrected for movement-related effects, spatially normalized to an EPI template (Montreal Neurological Institute reference brain<sup>29</sup>), an approximation of canonical space<sup>30</sup>, resampled into 4-mm cubic voxels, and finally spatially smoothed (8-mm gaussian kernel).

For the statistical model an event-related design matrix including all conditions of

interest was specified using the canonical haemodynamic response function for all event types<sup>28</sup>. The data was high-pass filtered, smoothed temporally with a 4-s full-width, half-maximum gaussian kernel, and rescaled to the global mean. Subject-specific contrasts were estimated using a fixed-effects model. Those contrast images were used to obtain subject-specific estimates for each effect. For group analysis, these estimates were entered in a second-level analysis treating subjects as a random effect, using a two-sample  $t$ -test at each voxel. Unless mentioned otherwise, resulting SPM-T maps were thresholded at  $P < 0.001$ , uncorrected for multiple comparisons, and only activations involving contiguous clusters of at least ten voxels are reported. Maxima and all coordinates are reported in MNI coordinates.

Received 25 October 2001; accepted 2 January 2002.

1. Penfield, W. & Roberts, L. *Speech and Brain Mechanisms* (Princeton Univ. Press, Princeton, New Jersey, 1959).
2. Demonet, J. F. *et al.* The anatomy of phonological and semantic processing in normal subjects. *Brain* **115**, 1753–1768 (1992).
3. Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. Lateralisation of phonetic and pitch discrimination in speech processing. *Science* **256**, 846–849 (1992).
4. Paulesu, E., Frith, C. & Frackowiak, R. The neural correlates of the verbal component of working memory. *Nature* **362**, 342–345 (1993).
5. Petersen, S., Fox, P., Posner, M., Mintun, M. & Raichle, M. Positron emission tomographic studies of the processing of single words. *J. Cogn. Neurosci.* **1**, 153–170 (1989).
6. Coltheart, M., Curtis, B., Atkins, P. & Haller, M. Models of reading aloud: dual-route and parallel-distributed processing approaches. *Psychol. Rev.* **100**, 589–608 (1993).
7. Chee, M. W. L. *et al.* Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron* **23**, 127–137 (1999).
8. Perani, D. *et al.* The bilingual brain—proficiency and age of acquisition of the second language. *Brain* **121**, 1841–1852 (1998).
9. Scarborough, D. L., Gerard, L. & Cortese, C. Independence of lexical access in bilingual word recognition. *J. Verb. Learn. Verb. Behav.* **23**, 84–99 (1984).
10. Macnamara, J. & Kushnir, S. Linguistic independence of bilinguals: the input switch. *J. Verb. Learn. Verb. Behav.* **10**, 480–487 (1971).
11. Gerad, L. D. & Scarborough, D. L. Language-specific lexical access of homographs by bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* **15**, 305–313 (1989).
12. Dijkstra, T., Timmermans, M. & Schriefers, H. On being blinded by your other language: effects of task demands on interlingual homograph recognition. *J. Mem. Lang.* **42**, 445–464 (2000).
13. Comrie, B. *The World's Major Languages* (Croom Helm, London, 1987).
14. Schmitt, B. M., Münte, T. F. & Kutas, M. Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* **37**, 473–484 (2000).
15. Kutas, M. & Hillyard, S. A. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* **207**, 203–205 (1980).
16. Rugg, M. D. Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Mem. Cogn.* **18**, 367–379 (1990).
17. Nobre, A. & McCarthy, G. Language-related fields potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *J. Neurosci.* **15**, 1090–1098 (1995).
18. Luck, S. J., Vogel, E. K. & Shapiro, K. L. Word meanings can be accessed but not reported during the attentional blink. *Nature* **383**, 616–618 (1996).
19. Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature* **331**, 585–589 (1988).
20. Xu, B. *et al.* Conjoint and extended neural networks for the computation of speech codes: the neural basis of selective impairment in reading words and pseudowords. *Cerebral Cortex* **11**, 267–277 (2001).
21. Bunge, S. A., Ochsner, K. N., Desmond, J. E., Glover, G. H. & Gabrieli, J. D. Prefrontal regions involved in keeping information in and out of mind. *Brain* **124**, 2074–2086 (2001).
22. Thompson-Schill, S. L., D'Esposito, M. & Kan, I. P. Effects of repetition and competition on activity in the left prefrontal cortex during word generation. *Neuron* **23**, 513–522 (1999).
23. MacDonald, A. W., Cohen, J. D., Stenger, V. A. & Carter, C. S. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* **288**, 1835–1838 (2000).
24. Weber-Fox, C. & Neville, H. Maturation constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *J. Cogn. Neurosci.* **8**, 231–256 (1997).
25. Alameda, J. R. & Cuetos, F. *Diccionario de Frecuencias de las Unidades Lingüísticas del Castellano* (Servicio de Publicaciones Universidad de Oviedo, Oviedo, 1995).
26. Rafel i Fontanals, J. *Diccionari de Freqüències* (Institut d'Estudis Catalans, Barcelona, 1998).
27. Winer, B. J., Brown, D. R. & Michels, K. M. *Statistical Principles in Experimental Designs* (McGraw-Hill, New York, 1991).
28. Friston, K. J. *et al.* Event-related fMRI: characterizing differential responses. *Neuroimage* **7**, 30–40 (1998).
29. Cocosco, C. A., Kollokian, V., Kwan, R. K. S. & Evans, A. C. BrainWeb: online interface to a 3D MRI simulated brain database. *Neuroimage* **5**, S425 (1997).
30. Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical, New York, 1988).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

### Acknowledgements

This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Spanish government.

### Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to T.F.M. (e-mail: [thomas.muente@medizin.uni-magdeburg.de](mailto:thomas.muente@medizin.uni-magdeburg.de)).

# Functional neurogenesis in the adult hippocampus

Henriette van Praag<sup>\*†</sup>, Alejandro F. Schinder<sup>\*†‡</sup>, Brian R. Christie<sup>\*†‡</sup>, Nicolas Toni<sup>\*†</sup>, Theo D. Palmer<sup>\*†‡</sup> & Fred H. Gage<sup>\*</sup>

<sup>\*</sup> Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

<sup>†</sup> These authors contributed equally to this work

There is extensive evidence indicating that new neurons are generated in the dentate gyrus of the adult mammalian hippocampus, a region of the brain that is important for learning and memory<sup>1–5</sup>. However, it is not known whether these new neurons become functional, as the methods used to study adult neurogenesis are limited to fixed tissue. We use here a retroviral vector expressing green fluorescent protein that only labels dividing cells, and that can be visualized in live hippocampal slices. We report that newly generated cells in the adult mouse hippocampus have neuronal morphology and can display passive membrane properties, action potentials and functional synaptic inputs similar to those found in mature dentate granule cells. Our findings demonstrate that newly generated cells mature into functional neurons in the adult mammalian brain.

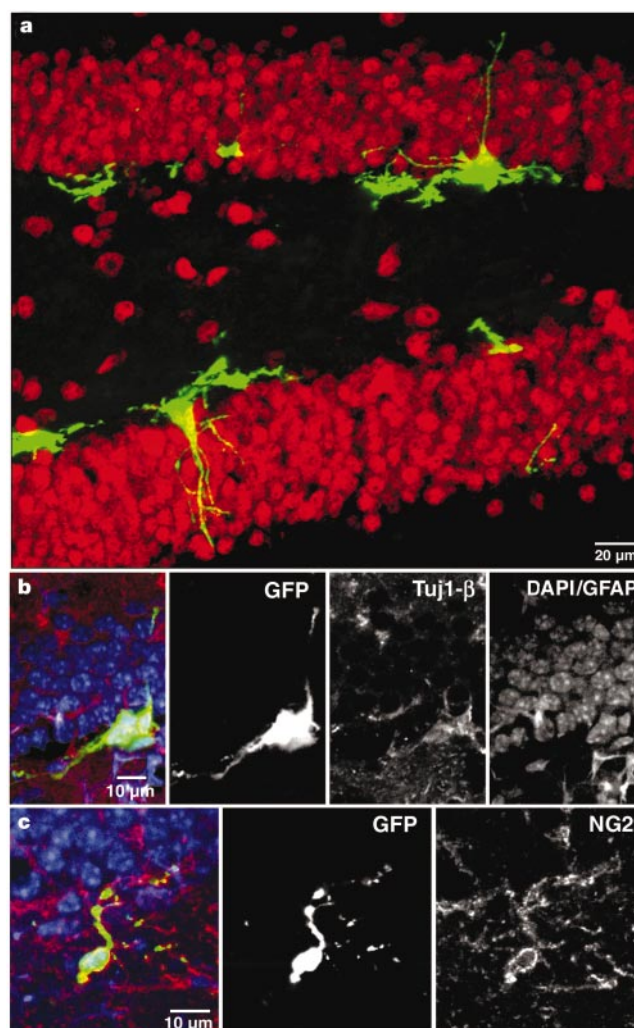
Hippocampal neurogenesis has been observed in adult animals from birds to humans<sup>1–6</sup>. The newly generated cells may have a function in cognition and brain repair. For example, manipulations that increase neurogenesis, such as an enriched environment<sup>7</sup> and exercise<sup>8</sup>, are associated with improved memory function and enhanced synaptic plasticity<sup>9</sup>. In addition, new neurons are generated in the hippocampus after stroke<sup>10</sup> and seizures<sup>11</sup>, and in the cortex after a selective lesion<sup>12</sup>, suggesting that they may be involved in recovery from injury. Stress, on the other hand, has been related to decreased cell proliferation and memory impairment<sup>13,14</sup>. However, the conclusions from all of these studies are based on correlations. It is still unclear whether newly generated cells that express neuronal markers become functional neurons.

Embryonic hippocampal progenitors can develop neuronal electrophysiological properties and are able to form functional synapses in culture<sup>15</sup>. Furthermore, adult hippocampal progenitors display voltage-dependent currents *in vitro*<sup>16,17</sup>. It remains unknown whether newly generated cells become functional neurons *in vivo*, mainly owing to the methods currently available to study neurogenesis. Both tritiated thymidine<sup>1,2,6</sup> and 5-bromodeoxyuridine (BrdU)<sup>3–5,7–9</sup> are used to label dividing cells in the adult brain; however, both methods require processing of the tissue and only label the soma. To overcome these limitations, we used a retroviral vector expressing green fluorescent protein (GFP), which only labels dividing cells<sup>18</sup>. GFP fills the soma as well as the processes, making structural analysis possible. Moreover, GFP can be visualized under fluorescence microscopy, allowing functional characterization of newly generated cells in hippocampal slices. Using this methodology, we have identified one-month-old neurons with functional properties similar to those of mature dentate granule cells in the adult hippocampus. We also show that newly generated neurons are initially smaller and reach a more mature morphology after 4 months.

To label proliferating cells in the adult hippocampus, a highly concentrated retroviral stock (1.5  $\mu$ l of  $5 \times 10^8$  infectious units per ml) carrying the GFP transgene was injected into the dentate gyrus.

One group of mice was housed with a running wheel<sup>8</sup> to enhance cell division before injection. Mice were killed after 48 h to assess cell phenotype shortly after proliferation, or after 4 weeks to determine cell fate. GFP-expressing (GFP<sup>+</sup>) cells were found in the dentate gyrus after both 48 h and 4 weeks. GFP<sup>+</sup> cell counts revealed that, in contrast to BrdU, retroviral labelling is quite variable and, therefore, unsuitable for absolute quantitative comparisons between controls and experimental mice (hereafter referred to as runners) (see Methods). In spite of this caveat, a relatively high percentage of GFP<sup>+</sup> cells expressed neuronal markers at 4 weeks in runners, making a detailed morphological and electrophysiological characterization of newly generated neurons possible (Table 1).

The phenotypes of the GFP<sup>+</sup> cells were examined by immunofluorescent labelling for neuronal and glial markers. At 48 h after viral injection, GFP<sup>+</sup> cells were found to be immature neurons (Tuj1- $\beta$ ), neural precursors (NG2) or glia (GFAP), but not mature neurons (NeuN or calbindin; Fig. 1 and Table 1). About 50% of the GFP<sup>+</sup> cells remained unidentified. A portion of these cells may be associated with the vasculature<sup>19</sup>. Most GFP<sup>+</sup> cells were located in the inner granule cell layers of the dentate gyrus, suggesting that the new cells arise from progenitors in the subgranular zone<sup>2</sup>.



**Figure 1** Dividing cells in the adult mouse dentate gyrus express early neural markers at 48 h after virus injection. **a**, Confocal micrograph (a merged image of 15 1- $\mu$ m optical sections) of GFP expression in the dentate gyrus in a section that was labelled for the neuronal marker NeuN (red). No co-labelling of GFP<sup>+</sup> cells with NeuN was observed. **b**, Single confocal plane of a cluster of GFP<sup>+</sup> cells labelled with the early neuronal marker Tuj1- $\beta$  (red). **c**, Single confocal plane of a GFP<sup>+</sup> cell with immunoreactivity for the progenitor marker NG2 (red). In **b** and **c**, nuclei (DAPI) and GFAP are blue.

<sup>‡</sup> Present addresses: Instituto de Investigaciones Bioquímicas Fundación Campomar, Av. Patricia Argentinas 435, (1405) Buenos Aires, Argentina (A.F.S.); Department of Psychology and Neuroscience Program, University of British Columbia Vancouver, British Columbia, V6T 1Z4, Canada (B.R.C.); Department of Neurosurgery, Stanford University, California 94305, USA (T.D.P.).

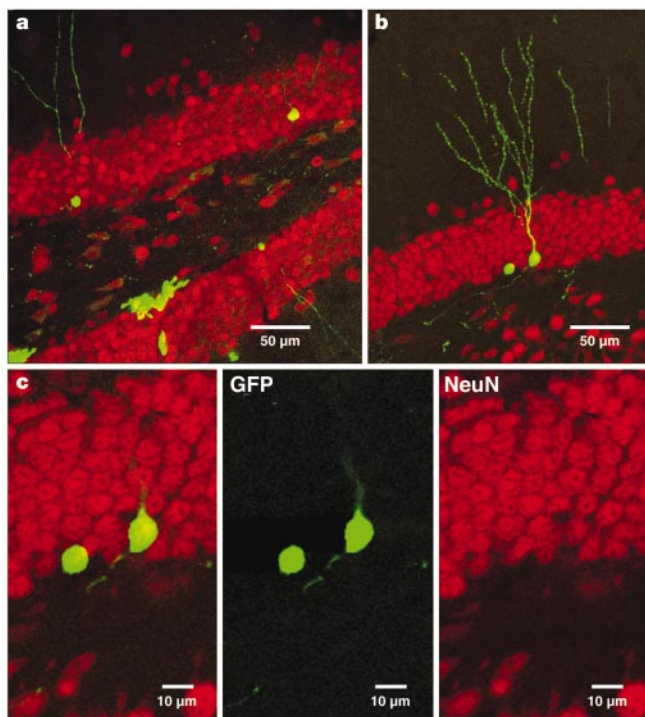


**Table 1 Number and phenotype of GFP<sup>+</sup> cells**

| Property of cells      | Control<br>48 h<br>( <i>n</i> = 5) | Running<br>48 h<br>( <i>n</i> = 5) | Control<br>4 weeks<br>( <i>n</i> = 5) | Running<br>4 weeks<br>( <i>n</i> = 8) |
|------------------------|------------------------------------|------------------------------------|---------------------------------------|---------------------------------------|
| GFP <sup>+</sup> cells | 265 (66)                           | 260 (45)                           | 87 (22)                               | 130 (38)                              |
| Spread (mm)            | 1.54 (0.16)                        | 1.58 (0.16)                        | 1.82 (0.28)                           | 1.47 (0.21)                           |
| NeuN (%)               | 0                                  | 0                                  | 2.2 (1.3)                             | 27.4 (9.3)**                          |
| Calbindin (%)          | 0                                  | 0                                  | 2.0 (1.2)                             | 22.5 (9.3)**                          |
| Tuj1-β (%)             | 9.2 (1.1)                          | 28.4 (8.3)*                        | 15.6 (6.9)                            | 31.7 (3.6)*                           |
| NG2 (%)                | 26.0 (12.2)                        | 35.0 (2.3)*                        | 7.7 (5.0)                             | 13.7 (3.8)                            |
| GFAP (%)               | 4.1 (1.9)                          | 4.6 (3.6)                          | 10.3 (4.2)                            | 8.6 (1.5)                             |

The total number of GFP<sup>+</sup> cells from the dentate gyrus is shown at indicated time points after viral injection. At 48 h, runners had a greater percentage of GFP<sup>+</sup> cells labelled for Tuj1-β ( $P < 0.04$ ) and NG2 ( $P < 0.05$ ) than controls. At 4 weeks, the percentage of GFP<sup>+</sup> cells co-labelling with neuronal markers was significantly higher in runners than in controls (NeuN,  $P < 0.01$ ; calbindin,  $P < 0.01$ ; Tuj1-β,  $P < 0.04$ ). All data are presented as means, with s.e.m. in parentheses. Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.01$ .

GFP<sup>+</sup> cells expressing neuronal markers were found 4 weeks after injection (Table 1). Moreover, as GFP is expressed throughout the cytoplasm, dendritic processes extending towards and into the molecular layer and an axon projecting into the hilar area were observed (Fig. 2). Newly generated cells expressing neuronal markers were present throughout the granule cell layer (Figs 2a and 3a). The percentage of GFP<sup>+</sup> cells that co-labelled with neuronal markers was significantly higher in runners (approximately 25%) than in controls (approximately 2%), supporting our previous observation that exercise enhances neurogenesis<sup>8,9</sup>. However, in both groups the percentages of GFP<sup>+</sup> neurons were lower than in the previous BrdU studies. This is probably due to differences in mechanisms of action between retrovirus and BrdU, or to transgene downregulation and/or shutdown of GFP expression. Downregulation may depend on the viral integration site into the genome of the host cell<sup>20</sup>, and on the degree of differentiation of the cells. Indeed, GFP fluorescence was fainter in neuronal cells than in other cell types.



**Figure 2** GFP<sup>+</sup> cells in the dentate gyrus 4 weeks after virus injection express mature neuronal markers. **a**, Overview of GFP expression in the dentate gyrus in a section (a merged image of 27 1-μm optical sections) that was labelled for the neuronal marker NeuN (red). **b**, GFP<sup>+</sup> cell (a merged image of 18 1-μm optical sections) immunoreactive for NeuN. **c**, Single confocal plane of the same cell, showing co-localization of GFP and NeuN.

**Table 2 Morphological characteristics of GFP<sup>+</sup> newly generated granule cells**

| Morphological characteristic       | Four-week-old neurons | Four-month-old neurons | Statistical significance ( <i>P</i> ) |
|------------------------------------|-----------------------|------------------------|---------------------------------------|
| Area of soma* (μm <sup>2</sup> )   | 80.0 (5.6)            | 138 (14)               | 0.002                                 |
| Total dendritic length (μm)        | 359 (45)              | 552 (49)               | 0.01                                  |
| Number of branch points            | 4.1 (0.7)             | 6.8 (0.6)              | 0.01                                  |
| Spine density† (μm <sup>-1</sup> ) | 0.77 (0.04)           | 1.19 (0.07)            | 0.0002                                |

Data are presented as means, with s.e.m. in parentheses. The column on the right shows the statistical significance, calculated using a *t*-test. *n* = 8 for all characteristics, except spine density for 4-month-old neurons, where *n* = 5.

\* Maximal cross-sectional area obtained from stereological measurements.

† Measured over a total dendritic length of 607 μm in newly generated cells at 4 weeks and 638 μm at 4 months after virus injection. The distances between the cell body and the beginning of the measured dendritic segments were similar for the 4-week-old (71 ± 9 μm) and 4-month-old (57 ± 5 μm) neurons ( $P = 0.2$ ).

To determine whether newborn neurons receive synaptic input, we carried out synaptophysin immunohistochemistry. Figure 3a shows a GFP<sup>+</sup>, calbindin-positive cell with synaptophysin staining, consistent with previous studies<sup>21</sup>. Additional evidence for synaptic inputs is the presence of dendritic spines (Fig. 3b–e), suggesting that glutamate-containing terminals contact these dendrites<sup>22</sup>. Furthermore, ultrastructural analysis confirmed that newly generated granule cells do receive synaptic inputs<sup>2</sup>. The synapses of these newly generated granule cells express mature features such as a pool of presynaptic vesicles facing a postsynaptic density (Fig. 3f).

We have shown that newly generated cells can express mature neuronal markers and display the distinct morphology of a dentate granule neuron. To assess whether maturation continues, we compared the morphological characteristics of newly generated neurons at 4 weeks (Fig. 3d, e) and at 4 months (Fig. 3b, c) after retroviral injection. All measured parameters (size of soma, total dendritic length, dendritic branching and spine density) increased by about 60% over this period (Table 2). Thus, newly generated neurons in the adult grow and mature over several months. These findings are consistent with studies showing that spine density and dendritic complexity of dentate granule cells increase from birth well into adulthood<sup>23</sup>. Spine counts at 4 months were similar to those reported in the literature for mature neurons<sup>24</sup>. Thus, newborn neurons acquire the morphological dimensions of mature granule cells over several months. Of note, newly generated granule cells are functionally integrated into the circuitry by 4 weeks (see below).

To investigate whether newly generated neurons are functional, electrophysiological recordings of GFP<sup>+</sup> cells were made in acute hippocampal slices. Several GFP<sup>+</sup> cells showed neuronal properties (Table 3). In the example shown in Fig. 4, membrane potentials recorded in response to depolarizing current steps displayed

**Table 3 Electrophysiological properties of newly generated and mature dentate granule cells**

| Electrophysiological property | Newly generated granule cells | <i>n</i> | Mature granule cells | <i>n</i> | Statistical significance ( <i>P</i> ) |
|-------------------------------|-------------------------------|----------|----------------------|----------|---------------------------------------|
| Resting potential (mV)        | −69.7 (1.7)                   | 9        | −74.8 (1.1)          | 27       | 0.06                                  |
| Input resistance (MΩ)         | 350 (37)                      | 11       | 388 (37)             | 25       | 0.62                                  |
| Time constant (ms)            | 16.6 (4.0)                    | 11       | 33.7 (3.7)           | 25       | 0.003                                 |
| Membrane capacitance* (pF)    | 42.3 (10.2)                   | 11       | 99.2 (12.7)          | 25       | 0.003                                 |
| Spiking threshold (mV)        | −45.8 (2.9)                   | 13       | −39.6 (1.5)          | 27       | 0.07                                  |
| Firing rate† (Hz)             | 27.5 (2.6)                    | 10       | 29.9 (4.8)           | 14       | 0.65                                  |
| Spontaneous activity‡ (Hz)    | 1.6 (0.8)                     | 6        | 0.80 (0.19)          | 9        | 0.33                                  |

Experiments on newly generated neurons (GFP<sup>+</sup>) were carried out 4–8 weeks after viral injection. Data are presented as means, with s.e.m. in parentheses. The column on the right shows the statistical significance, calculated using a *t*-test.

\* Calculated from the input resistance and the time constant, as  $C_m = \tau_m/R_{in}$ .

† Measured in response to a 100-pA depolarizing current step.

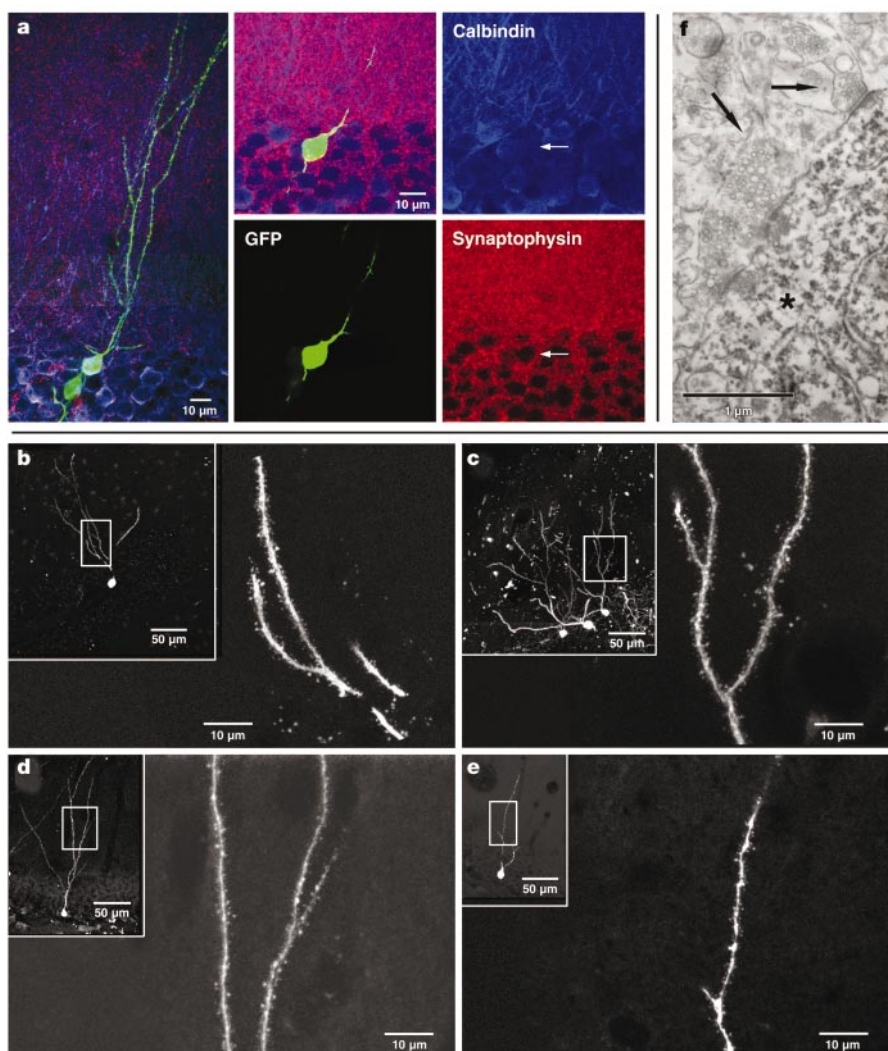
‡ Spontaneous postsynaptic responses measured in voltage or current clamp.

repetitive spiking with increasing frequency, reaching a plateau at about 140 Hz (Fig. 4b, d). Action potentials displayed a threshold ( $V_{th}$ ) of  $-43$  mV, conspicuous after hyperpolarization and frequency adaptation (Fig. 4b–d), similar to mature dentate granule cells under comparable recording conditions<sup>25,26</sup> (Fig. 5 and Table 3). To assess whether this cell had functional inputs, spontaneous postsynaptic currents were recorded (Fig. 4e). Currents displayed fast onset (less than 1 ms) and a slower exponential decay (greater than 5 ms), typical of postsynaptic responses to fast neurotransmitters, such as glutamate and  $\gamma$ -aminobutyric acid (GABA). Similar results were obtained in five additional experiments, where spontaneous postsynaptic responses were observed at a frequency of 0.4–5 Hz and with peak amplitudes ranging from 0.5 to 1 mV (under current clamp) and 10 to 20 pA (under voltage clamp; see Table 3). After fixation of the slice, co-labelling for GFP, Alexa Fluor 594 and NeuN showed that this cell had electrophysiological and morphological properties of a functional granule cell (Fig. 4f, g).

The main excitatory input to dentate granule cells is the perforant pathway. To assess whether newly generated neurons receive appropriate projections, postsynaptic currents were recorded in response to extracellular stimulation of the perforant path. Figure 5 depicts

recordings from a GFP<sup>+</sup> cell and an adjacent mature granule cell with similar spiking properties. In both cells, excitatory postsynaptic currents were evoked in response to paired-pulse stimulation (Fig. 5c). Interestingly, although the same afferents were stimulated, the mature neuron displayed paired-pulse facilitation whereas the newly generated neuron showed depression, suggesting differences in presynaptic release properties. The observation that newly generated granule cells receive inputs from the perforant path ( $n = 4$ ) strongly suggests that they are functionally integrated in the hippocampus.

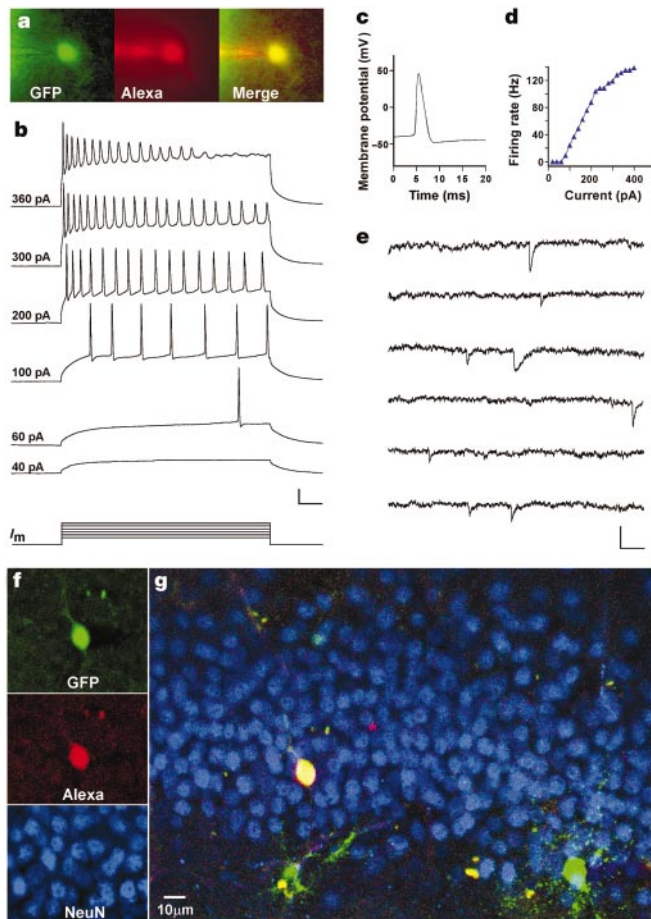
A total of 13 GFP<sup>+</sup> cells with neuronal properties were characterized from 13 mice (Table 3). For comparison, recordings were carried out from 27 mature dentate granule cells obtained from 23 mice (Fig. 5 and Table 3). Most properties (resting potential ( $V_r$ ), input resistance ( $R_{in}$ ), threshold potential for spiking and firing rate) were found to be similar between the two groups. In addition, spontaneous postsynaptic responses (Fig. 4e and Table 3) displayed similar characteristics. Together with the observation that newly generated neurons receive inputs from the perforant path, these findings show a remarkable functional similarity between newly generated and mature granule cells.



**Figure 3** Newly generated neurons receive synaptic inputs. **a**, Confocal photomicrograph (a merged image of 10 0.5- $\mu$ m optical sections) of a GFP<sup>+</sup> cell, immunolabelled for calbindin (blue) and synaptophysin (red). Adjacent panels (merged images of 4 0.5- $\mu$ m optical sections) show co-localization with calbindin and synaptophysin. **b–e**, Visualization of spines in GFP<sup>+</sup> cells (merged images of 12–33 optical sections of

0.5  $\mu$ m). The boxed areas in the insets correspond to the enlarged images of the dendrites. Cells were analysed 4 months (**b, c**) or 4 weeks (**d, e**) after virus injection. **f**, Electron micrograph of synaptic terminals (arrows) on the soma of a GFP<sup>+</sup> neuron (asterisk) in the granule cell layer.



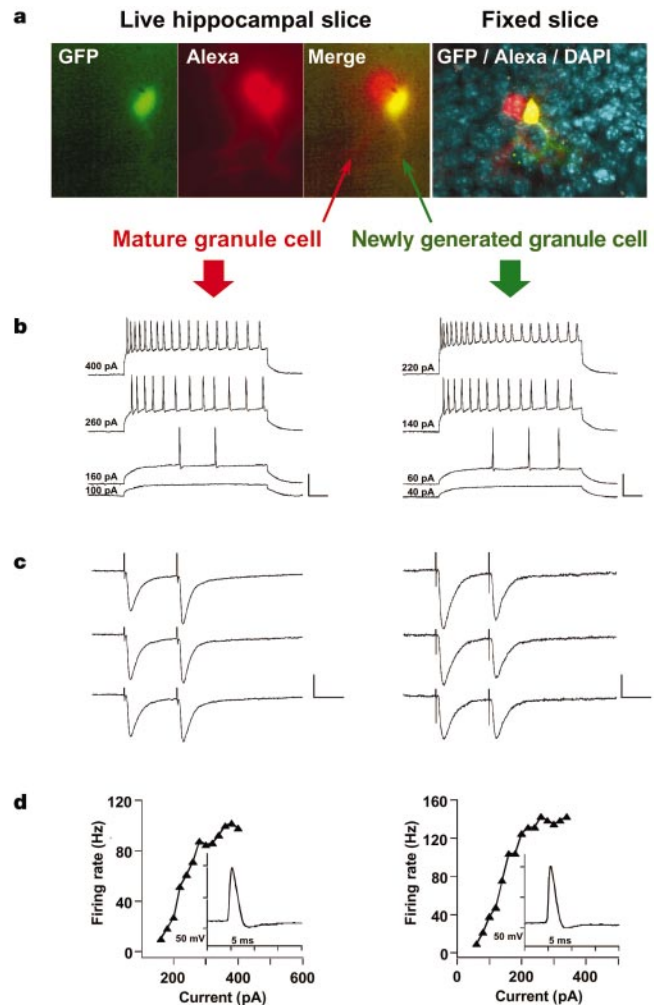


**Figure 4** Newly generated cells display neuronal electrophysiological properties. **a**, Fluorescent micrographs of a GFP<sup>+</sup> cell filled with Alexa through a recording pipette. **b**, Membrane potential in response to depolarizing currents ( $I_m$ ; 500 ms, 20–400 pA) recorded under current clamp at the resting potential (−73.5 mV). Numbers on the left indicate stimulus size. Scale: 25 mV, 50 ms. **c**, Action potential recorded at 80 pA. **d**, Firing rate versus current curve. **e**, Spontaneous postsynaptic currents recorded under voltage clamp (−80 mV). Scale: 20 pA, 100 ms. **f**, Confocal micrographs taken from a single plane (1  $\mu$ m) after fixation of the slice. **g**, Overview of the dorsal blade of the granule cell layer (merged z-series of 23 planes).

The analysis of passive membrane properties ( $V_r$ ,  $R_{in}$  and the time constant  $\tau_m$ ) revealed that newly generated neurons have a significantly faster  $\tau_m$  compared with mature dentate granule cells (Table 3). The membrane capacitance ( $C_m$ ), calculated as  $C_m = \tau_m/R_{in}$ , was found to be smaller in newly generated than in mature granule neurons, indicating that 4-week-old neurons are smaller than mature dentate granule cells<sup>27</sup> and confirming our own morphological observations (Fig. 5 and Table 2).

Recordings were performed from a total of 58 GFP<sup>+</sup> cells with different morphology and location in the dentate gyrus. All cells that displayed neuronal properties ( $n = 13$ ) had a round soma located in the granule cell layer, dendrites extending towards the molecular layer, and, in general, low fluorescence intensity (Figs 4f and 5a). The additional 45 cells were non-excitable, had a markedly low  $R_{in}$  ( $209 \pm 31$  M $\Omega$ ) and  $\tau_m$  ( $9.8 \pm 16$  ms), and did not receive functional inputs. Typically, these cells had a large and irregular soma, multiple branches, and brighter fluorescence (data not shown). It is likely that these non-spiking cells represent a mixed population of neural precursors.

Our results demonstrate that the adult mammalian dentate gyrus gives rise to new functional neurons that are integrated into the hippocampal circuitry. It will be crucial to determine the



**Figure 5** Newly generated neurons are functionally similar to mature dentate granule cells. Consecutive recordings of a GFP<sup>+</sup> neuron (right) and an adjacent granule cell (left) are shown. The GFP<sup>+</sup> neuron ( $C_m = 32$  pF) was smaller and, therefore, younger than the unlabelled neuron ( $C_m = 60$  pF). **a**, Fluorescence micrographs of both cells filled with Alexa (left panels). The right panel is a confocal micrograph taken after fixation, with GFP (green), Alexa (red) and DAPI (blue) visualized. **b**, Membrane potential in response to depolarizing current steps. Stimulus size is indicated on the left. Scale: 50 mV, 50 ms. **c**, Postsynaptic currents evoked by paired-pulse stimulation (100  $\mu$ A, 0.2 ms, with a 50-ms interval) of the perforant path. The position of the stimulating electrode was identical for both neurons. Transients before the traces are stimulus artifacts. Scale: 100 pA, 25 ms (left); 40 pA, 25 ms (right). **d**, Spiking rate versus current curve. Insets show a single action potential.

functional significance of these newly generated cells in the adult brain. One possible hypothesis is that new neurons may be required to replace dying neurons<sup>28</sup>. Another possibility is that young neurons provide a greater degree of plasticity to the mature brain. This enhanced plasticity would become apparent from the integration of new functional units whose connectivity may be shaped by experience. □

## Methods

### Subjects and stereotaxic surgery

Female C57Bl/6 mice, 4–5 weeks old (Jackson Laboratories), were housed in standard conditions ( $n = 10$ ) or with a running wheel ( $n = 80$ ), four or five mice per cage. After one week, mice were anaesthetized (100  $\mu$ g ketamine, 10  $\mu$ g xylazine in 10  $\mu$ l saline per gram) and virus (1.5  $\mu$ l at  $0.32 \mu$ l min<sup>−1</sup>) was infused into the right dentate gyrus (anteroposterior = −2 mm from bregma; lateral = 1.5 mm; ventral = 2 mm). For electrophysiology ( $n = 53$ ), another injection of virus (1.5  $\mu$ l) was given in the same dentate gyrus (anteroposterior = −3 mm from bregma; lateral = 3 mm; ventral = 3 mm). Mice were killed 48 h,



4 weeks or 4 months after surgery for morphological experiments, and 4–8 weeks after injections for electrophysiology.

## Immunocytochemistry

We carried out triple labelling on 40- $\mu\text{m}$  free-floating coronal sections as described<sup>3,19,21</sup>. We applied antibodies in TBS with 3% donkey serum and 0.3% Triton X-100. Antibodies for NeuN (mouse monoclonal, 1:10), NG2 (rabbit polyclonal, Chemicon, 1:500) and GFAP (glialfibrillary acidic protein, guinea-pig polyclonal, Advanced Immunochemicals, 1:500) were combined. In additional series of sections, Tuj1- $\beta$  (mouse monoclonal, ABCO, 1:500), calbindin (rabbit polyclonal, S. Want, 1:500) and GFAP were pooled. Synaptophysin (mouse monoclonal, Boehringer-Mannheim, 1:50), calbindin and GFAP were also combined. Corresponding secondary antibodies (donkey anti-mouse Cy3, donkey anti-rabbit Cy5 and donkey anti-guinea-pig AMCA, 1:250) were used. We visualized nuclei with 4'-6-diaminodino-2-phenylindole (DAPI). For BrdU labelling, mouse anti-BrdU (Boehringer-Mannheim, 1:400) was used with biotinylated donkey anti-mouse (1:250) as a secondary antibody and diaminobenzidine as a chromagen (Vector Laboratories).

## Cell quantification

Cells were counted using a confocal microscope (Bio-Rad) for GFP and a brightfield microscope (Leitz) for BrdU through a  $\times 40$  objective throughout the rostral–caudal extent of the granule cell layer in 1-in-6 series of 40- $\mu\text{m}$  sections (240  $\mu\text{m}$  apart). The total number of cells in the dentate gyrus was calculated on the basis of the multiple of the number of cells counted per hippocampus, and the sampling frequency.

## Cell proliferation

To compare between effects of retrovirus and BrdU on cell proliferation, additional mice were used ( $n = 5$  for runners,  $n = 4$  for controls). After one week mice were given an intraperitoneal injection of BrdU (100  $\mu\text{g g}^{-1}$ ) and killed after 48 h. BrdU-positive cell number differed significantly between controls ( $964 \pm 222$ ) and runners ( $1,909 \pm 306$ ) ( $P < 0.05$ ), suggesting that BrdU is preferable to retrovirus for cell quantification.

## Spine density and soma measurements

Each GFP<sup>+</sup> neuron that we used for spine counts was imaged at  $\times 40$  to obtain an overview. Dendrites were imaged through a  $\times 100$  objective with a digital zoom of 2. We merged z-series of 12–33 optical sections of 0.5  $\mu\text{m}$  for the spine counts. Spines were defined by a visible neck and a spine head of at least 4 pixels. To quantify maximal cross-sectional somatic area, the total number of pixels was obtained using Adobe Photoshop. Measurements were carried out on merged z-series (12–18 1- $\mu\text{m}$  optical sections) and then converted to  $\mu\text{m}^2$ .

## Electron microscopy

GFP<sup>+</sup> cells in 100- $\mu\text{m}$  sections were injected under a fluorescence microscope with 0.4% fluorescein-conjugated horseradish peroxidase (HRP, Sigma). Slices were processed for HRP<sup>29</sup>, post-fixed in 1% osmium, and embedded in epoxy resin. Synapses were analysed on 35-nm thick serial sections on a jeol 100cxII electron microscope at 60 Kv, at a magnification of less than  $\times 10,000$ .

## Viral vector production

A retroviral vector based on the Moloney murine leukaemia virus expressing GFP was used. A stable NIT-GFP packaging cell line was generated as described previously<sup>30</sup>. Virus-containing supernatant was collected from clone 293gp/NIT-GFPc11 after transfection (Superfect, Qiagen) with pVSVG, and concentrated as described. Final virus titres were  $5 \times 10^8$  neomycin-resistant (*neo*<sup>r</sup>) colonies per ml of virus, as measured by G418-resistant colony formation on NIH 3T3 cells.

## Electrophysiology

Four to eight weeks after surgery, runner mice were anaesthetized and decapitated. Brains were removed into a chilled solution containing (in mM): 110 choline Cl<sup>−</sup>, 2.5 KCl, 1.3 NaH<sub>2</sub>PO<sub>4</sub>, 25.0 NaHCO<sub>3</sub>, 0.5 CaCl<sub>2</sub>, 7 MgCl<sub>2</sub>, 20 dextrose, 1.3 Na<sup>+</sup> ascorbate, 0.6 Na<sup>+</sup> pyruvate, 5.5 kynurenic acid. Slices (200–400  $\mu\text{m}$  thick) were cut as described previously<sup>9</sup>, transferred to a chamber containing artificial cerebrospinal fluid (in mM: 125.0 NaCl, 2.5 KCl, 1.3 NaH<sub>2</sub>PO<sub>4</sub>, 25.0 NaHCO<sub>3</sub>, 2 CaCl<sub>2</sub>, 1.3 MgCl<sub>2</sub>, 1.3 Na<sup>+</sup> ascorbate, 0.6 Na<sup>+</sup> pyruvate, 10 dextrose), bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub> (pH 7.4, 320 mOsm), and stored at 30 °C. Using fluorescein optics (Leica DMLFS), slices were scanned for GFP<sup>+</sup> cells in the dentate gyrus. Microelectrodes (6–8 M $\Omega$ ) were pulled from borosilicate glass (KG-33, Garner Glass) and filled with (in mM): 120.0 potassium gluconate, 20 KCl, 5 NaCl, 4 MgCl<sub>2</sub>, 0.1 EGTA, 10.0 HEPES, 4.0 Tris-ATP, 0.3 Tris-GTP, 10 phosphocreatine (pH 7.4, 295 mOsm).

Whole-cell recordings (Axopatch 200B, Axon Instruments) were filtered at 2 kHz, digitized and acquired onto a PC using Labview-based acquisition and analysis software (L. Campbell, Salk Institute). Series resistance was typically 10–30 M $\Omega$ . We delivered stimuli at 5-s intervals. Criteria to include cells in the analysis were: (1) co-labelling with Alexa Fluor 594 (10  $\mu\text{g ml}^{-1}$ , Molecular Probes) or visual confirmation of GFP in the pipette tip; (2) resting potential less than  $-50$  mV; and (3) leak current less than 300 pA. Firing rate was calculated as the inverse of the inter-spike interval between the first 2 spikes of the train. We calculated  $\tau_m$  by fitting a single exponential function to the membrane potential traces. Extracellular stimulation was done using a concentric bipolar electrode

(50- $\mu\text{m}$  diameter; FHC) and a stimulus isolation amplifier (Getting Instruments). After recording, slices were stored in 4% paraformaldehyde for over 24 h and processed for NeuN, followed by confocal analysis. We used GFP-negative dentate granule cells of slices from mice injected with virus as controls.

Received 6 November 2001; accepted 2 January 2002.

- Altman, J. & Das, G. D. Autoradiographic and histological evidence of postnatal neurogenesis in rats. *J. Comp. Neurol.* **124**, 319–335 (1965).
- Kaplan, M. S. & Hinds, J. W. Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs. *Science* **197**, 1092–1094 (1977).
- Kuhn, H. G., Dickinson-Anson, H. & Gage, F. H. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J. Neurosci.* **16**, 2027–2033 (1996).
- Kornack, D. R. & Rakic, P. Continuation of neurogenesis in the hippocampus of the adult macaque monkey. *Proc. Natl Acad. Sci. USA* **96**, 5768–5773 (1999).
- Eriksson, P. S. *et al.* Neurogenesis in the adult human hippocampus. *Nature Med.* **4**, 1313–1317 (1998).
- Barnea, A. & Nottebohm, F. Seasonal recruitment of hippocampal neurons in adult free-ranging black-capped chickadees. *Proc. Natl Acad. Sci. USA* **91**, 11217–11221 (1994).
- Kempermann, G., Kuhn, H. G. & Gage, F. H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493–495 (1997).
- van Praag, H., Kempermann, G. & Gage, F. H. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nature Neurosci.* **2**, 266–270 (1999).
- van Praag, H., Christie, B. R., Sejnowski, T. J. & Gage, F. H. Running enhances neurogenesis, learning and long-term potentiation in mice. *Proc. Natl Acad. Sci. USA* **96**, 13427–13431 (1999).
- Liu, J., Solway, K., Messing, R. O. & Sharp, F. R. Increased neurogenesis in the dentate gyrus after transient global ischemia in gerbils. *J. Neurosci.* **18**, 7768–7778 (1998).
- Parent, J. M. *et al.* Dentate granule neurogenesis is increased by seizures and contributes to aberrant network reorganization in the adult hippocampus. *J. Neurosci.* **17**, 3727–3738 (1997).
- Magavi, S. S., Leavitt, B. R. & Macklis, J. D. Induction of neurogenesis in the neocortex of adult mice. *Nature* **405**, 951–955 (2000).
- Gould, E., Cameron, H. A., Daniels, D. C., Wooley, C. S. & McEwen, B. S. Adrenal hormones suppress cell division in the adult rat dentate gyrus. *J. Neurosci.* **12**, 3642–3650 (1992).
- Lemaire, V., Koehl, M., Le Moal, M. & Abrous, D. N. Prenatal stress produces learning deficits associated with an inhibition of neurogenesis in the hippocampus. *Proc. Natl Acad. Sci. USA* **97**, 11032–11037 (2000).
- Vicario-Abejon, C., Collin, C., Tsoulfas, P. & McKay, R. D. Hippocampal stem cells differentiate into excitatory and inhibitory neurons. *Eur. J. Neurosci.* **12**, 677–688 (2000).
- Sah, D. W., Ray, J. & Gage, F. H. Regulation of voltage- and ligand-gated currents in rat hippocampal progenitor cells *in vitro*. *J. Neurobiol.* **32**, 95–110 (1997).
- Roy, N. S. *et al.* *In vitro* neurogenesis by progenitor cells isolated from the adult hippocampus. *Nature Med.* **6**, 271–277 (2000).
- Lewis, P. F. & Emerman, M. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J. Virol.* **68**, 510–516 (1994).
- Palmer, T. D., Willhoite, A. R. & Gage, F. H. Vascular niche for adult hippocampal neurogenesis. *J. Comp. Neurol.* **425**, 479–494 (2000).
- Xu, L., Yee, J. K., Wolff, J. A. & Friedmann, T. Factors affecting long-term stability of Moloney murine leukemia virus-based vectors. *Virology* **171**, 331–341 (1989).
- Markakis, E. A. & Gage, F. H. Adult-generated neurons in the dentate gyrus send axonal projections to field CA3 and are surrounded by synaptic vesicles. *J. Comp. Neurol.* **406**, 449–460 (1999).
- Edwards, F. A. Anatomy and electrophysiology of fast central synapses lead to a structural model for long-term potentiation. *Physiol. Rev.* **75**, 759–787 (1995).
- Crain, B., Cotman, C., Taylor, D. & Lynch, G. A quantitative electron microscopic study of synaptogenesis in the dentate gyrus of the rat. *Brain Res.* **63**, 195–204 (1973).
- Suzuki, F., Makiura, Y., Guilhem, D., Sorensen, J. C. & Onteniente, B. Correlated axonal sprouting and dendritic spine formation during kainate-induced neuronal morphogenesis in the dentate gyrus of adult mice. *Exp. Neurol.* **145**, 203–213 (1997).
- Staley, K. J., Otis, T. S. & Mody, I. Membrane properties of dentate gyrus granule cells: comparison of sharp microelectrode and whole-cell recordings. *J. Neurophysiol.* **67**, 1346–1358 (1992).
- Lübke, J., Frotscher, M. & Spruston, N. Specialized electrophysiological properties of anatomically identified neurons in the hilar region of the rat fascia dentate. *J. Neurophysiol.* **79**, 1518–1534 (1998).
- Coleman, P. A. & Miller, R. F. Measurement of passive membrane parameters with whole-cell recording from neurons in the intact amphibian retina. *J. Neurophysiol.* **61**, 218–230 (1989).
- Biehl, M., Cooper, C. M., Winkler, J. & Kuhn, H. G. Analysis of neurogenesis and programmed cell death reveals a self-renewing capacity in the adult rat brain. *Neurosci. Lett.* **291**, 17–20 (2000).
- Coleman, L. A. & Friedlander, M. J. Intracellular injections of permanent tracers in the fixed slice: a comparison of HRP and biocytin. *J. Neurosci. Methods* **44**, 167–177 (1992).
- Palmer, T. D., Markakis, E. A., Willhoite, A. R., Safar, F. & Gage, F. H. Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS. *J. Neurosci.* **19**, 8487–8497 (1999).

## Acknowledgements

We are grateful for the assistance of L. R. Kitabayashi and J. Simon in preparing the figures. We thank E. Bushong for assistance with electron microscopy, and N. Manison and R. Summers for help with confocal imaging. We also thank D. Cuizon, S. Forbes, M. Matthews, B. Miller and L. Moore for technical help, and M. L. Gage and K. Ganguly for comments on the manuscript. N.T. is supported by the Swiss Research Foundation. This work was supported by grants from the National Institutes of Health, Christopher Reeve Paralysis Foundation, and The Lookout Fund.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence should be addressed to H.v.P.  
(e-mail: vanpraag@salk.edu).

# Monoclonal mice generated by nuclear transfer from mature B and T donor cells

Konrad Hochedlinger\*‡ & Rudolf Jaenisch\*†

\* Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

† Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

‡ University of Vienna, Dr Karl Lueger Ring 1, A-1010 Vienna, Austria

**Cloning from somatic cells is inefficient, with most clones dying during gestation<sup>1,2</sup>. Cloning from embryonic stem (ES) cells is much more effective, suggesting that the nucleus of an embryonic cell is easier to reprogram<sup>3–7</sup>. It is thus possible that most surviving clones are, in fact, derived from the nuclei of rare somatic stem cells present in adult tissues rather than from the nuclei of differentiated cells as has been assumed<sup>1,8,9</sup>. Here we report the generation of monoclonal mice by nuclear transfer from mature lymphocytes. In a modified two-step cloning procedure, we established ES cells from cloned blastocysts and injected them into tetraploid blastocysts to generate mice. In this approach, the embryo is derived from the ES cells and the extra-embryonic tissues from the tetraploid host<sup>6</sup>. Animals cloned from a B-cell nucleus were viable and carried fully rearranged immunoglobulin alleles in all tissues. Similarly, a mouse cloned from a T-cell nucleus carried rearranged T-cell-receptor genes in all tissues. This is an unequivocal demonstration that a terminally differentiated cell can be reprogrammed to produce an adult cloned animal.**

The question of whether nuclei from specialized cells retain the potential to direct development of an animal was raised more than 40 years ago in the seminal nuclear-transfer experiments with amphibians<sup>10–12</sup>. However, the differentiation state of the cell that produced cloned frogs or cloned mammals has not been unequivocally resolved because no stable markers have been used to unambiguously identify the donor nucleus that gave rise to the rare cloned animal<sup>13–16</sup>. To derive cloned animals from nuclei that are genetically marked by terminal differentiation, we performed nuclear transfer using peripheral lymphocytes as donors. During normal lymphocyte maturation in bone marrow and thymus, B and T cells undergo genomic rearrangements at the immunoglobulin and T cell receptor (TCR) loci, respectively, and migrate into the periphery as resting, functional cells until they encounter antigens<sup>17,18</sup>. Peripheral lymph nodes contain, therefore, an almost pure population of terminally differentiated B and T lymphocytes that are genetically marked by these intrinsic DNA rearrangements.

Previous attempts to derive mice from lymphocytes by direct transfer of cloned blastocysts into the uteri of recipient females were unsuccessful<sup>7</sup>. We, therefore, developed a two-step procedure with the derivation of ES cells from cloned blastocysts followed by tetraploid embryo complementation to generate monoclonal mice. This method involves the injection of ES cells into a tetraploid host blastocyst, where the placenta is derived from the tetraploid host cells and the embryo from the injected donor ES cells<sup>6,19</sup>. Using

this two-step approach instead of direct blastocyst transfer to derive cloned animals allows for the generation of multiple mice from a single cloned embryo that was competent to produce an ES cell line.

We transferred nuclei from peripheral lymph node cells of C57BL/6 × DBA/2 F<sub>1</sub> (cell line LN1) or 129/SvJae × C57BL/6 F<sub>1</sub> (LN2) mice into enucleated oocytes to derive cloned blastocysts and subsequently ES cells, as has been demonstrated previously<sup>20–22</sup>. The overall efficiency of this procedure was low: only 41 blastocysts and 2 ES cell lines were obtained from about 1,000 transplanted oocytes (Table 1). The efficiency was about ten times lower than that from other donor cell populations<sup>22</sup> and might be due to inefficient reprogramming of the lymphocyte genome or differences in the sensitivity of the lymphocyte nuclei to the nuclear transfer protocol. To determine the origin and differentiation state of the donor nucleus, Southern blot analysis was performed on genomic DNA isolated from the two ES cell lines using probes recognizing either the immunoglobulin or TCR loci (Fig. 1a). These loci undergo sequential genomic rearrangements during differentiation, with randomly chosen variable (V), diversity (D) and joining (J) segments being assembled to form a functional B or T cell receptor<sup>17,18</sup>. Both alleles of the immunoglobulin heavy chain (IgH) locus and at least one allele of the light chain (IgL) locus were rearranged in LN1 ES cells (Fig. 1a), demonstrating that this line was derived from a mature B-cell nucleus. Similarly, the LN2 line had two rearranged alleles at both the TCR-α and TCR-β loci, indicating its derivation from a mature T-cell nucleus (Fig. 1a).

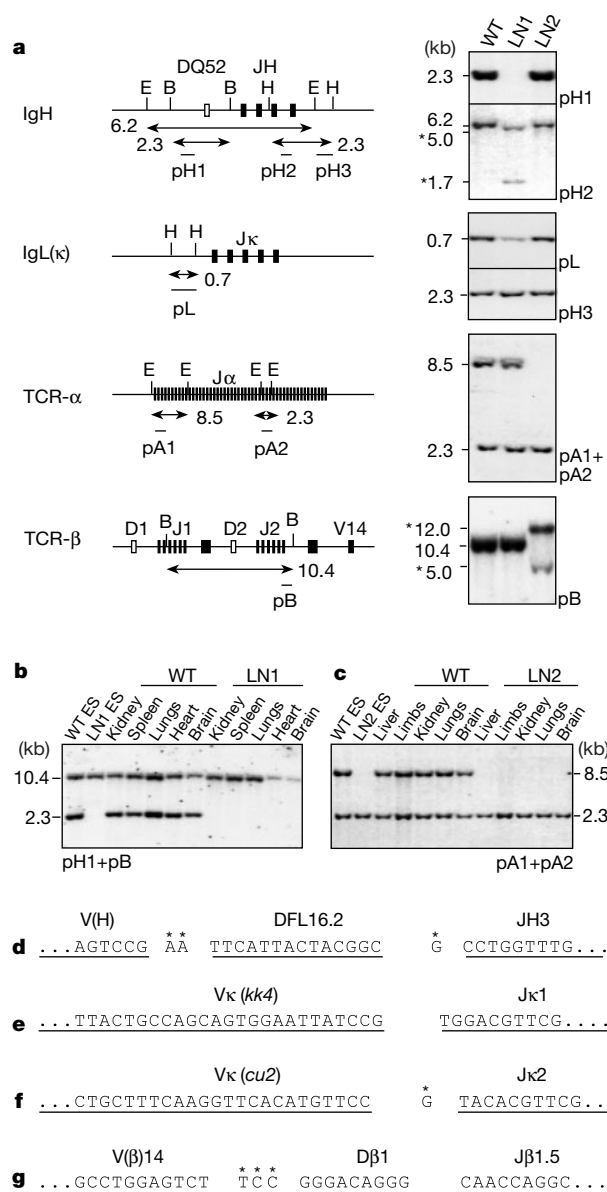
Using degenerate primers recognizing most of the variable elements of the IgH and IgL loci, we were able to amplify and sequence both the IgH and IgL rearrangements from genomic DNA of the LN1 line (Fig. 1d–f). Sequence comparison with the DnaPlot database (<http://www.dnaplot.de>) revealed that the functional heavy-chain region used a variable element of the VH6 family in combination with a DFL16.2 D element and a JH3 J element (Fig. 1d). The nonfunctional allele could not be detected by this polymerase chain reaction (PCR) assay. Analysis of the IgL locus revealed the presence of two rearranged alleles. One allele used a variable element of the Vκ4/5 group, subgroup IV, with similarity to the *kk4* gene, in combination with a Jκ1 element (Fig. 1e). The second allele used a variable element of the Vκ1 group, subgroup II, with similarity to the *cu2* gene, in combination with a Jκ2 element (Fig. 1f). These data, together with results from the Southern blot analyses, suggest that one κ allele had rearranged by deletional joining whereas the other allele had rearranged by inversional joining<sup>23</sup>. A PCR assay detecting all possible D-to-J rearrangements at the TCR-β locus of ES cell line LN2 showed that the active allele had recombined Dβ1 to one of the Jβ1 elements as assessed by the absence of a germline-specific band, whereas the inactive allele had recombined Dβ1 to Jβ2.5 with no further V-to-DJ rearrangement (data not shown).

Both lines were injected into tetraploid blastocysts to assess their potential for generating mice entirely derived from ES cells<sup>6,19</sup>. This procedure tests whether all the embryonic lineages of a mouse (giving rise to the animal) can be generated from ES cells. However, the procedure does not test whether extra-embryonic lineages (giving rise to the placenta) can be formed, because they are derived from the tetraploid host blastocyst. Nineteen mice were obtained from ES cell line LN1 (Table 2) with a frequency similar to that obtained from normal ES cell lines<sup>6</sup>. Southern blot analysis of

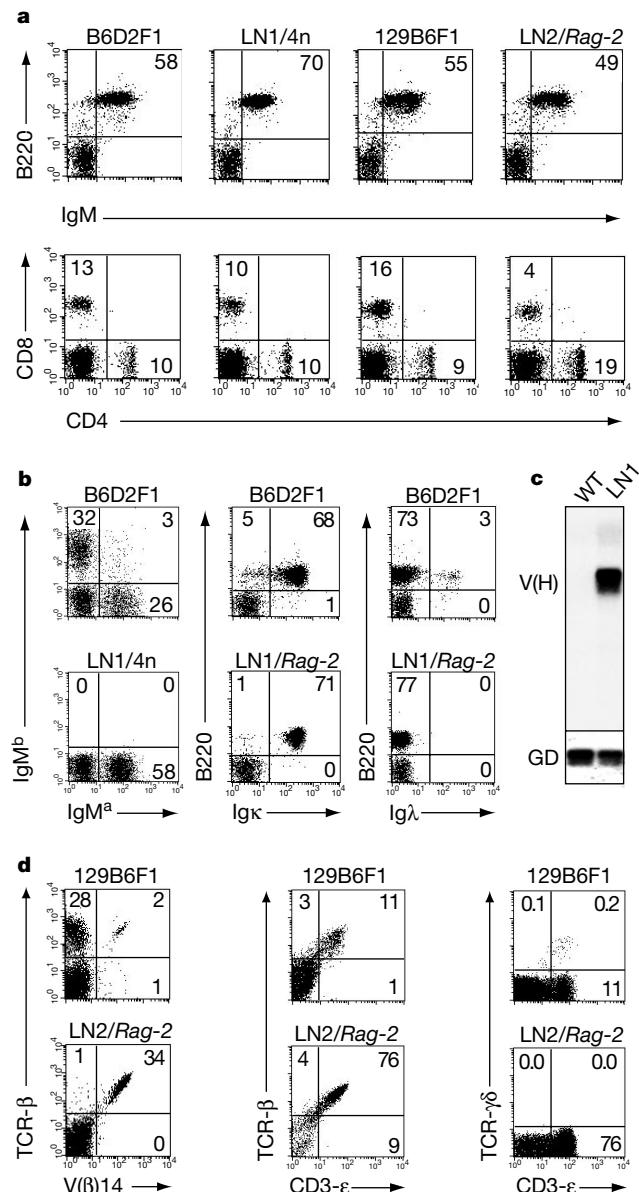
**Table 1** Preimplantation development of cloned embryos

| Donor tissue       | Eggs injected | Visible pronuclei | Two cells | Four cells | Morula | Blastocyst | ES cell line |
|--------------------|---------------|-------------------|-----------|------------|--------|------------|--------------|
| B6D2F1 lymph node  | 630           | 448               | 349       | 116        | 73     | 21         | 1 (LN1)      |
| 129B6F1 lymph node | 350           | 310               | 275       | 111        | 63     | 20         | 1 (LN2)      |
| Total              | 980           | 758               | 624       | 227        | 136    | 41         | 2            |

Only surviving eggs were counted. B6D2F1, C57BL/6 × DBA/2 F<sub>1</sub>; 129B6F1, 129/SvJae × C57BL/6 F<sub>1</sub>.



**Figure 1** Derivation of ES cells and monoclonal mice by nuclear transfer from mature B and T lymphocytes. **a**, Left: Schematic maps of the immunoglobulin and TCR loci including all relevant restriction sites, fragment lengths (in kilobases), and Southern blot probes (maps not drawn to scale). Characteristic elements are illustrated by boxes and labelled. B, *Bam*HI; E, *Eco*RI; H, *Hind*III. Right: Southern blot analyses of genomic DNA of ES cell lines LN1 and LN2 to detect rearrangements at the immunoglobulin and TCR loci, respectively. The following restriction enzyme/probe combinations were used: *Bam*HI/pH1 and *Eco*RI/pH2 for the IgH locus; *Hind*III/pL and *Hind*III/pH3 (loading control) for the IgL locus; *Eco*RI/pA1+pA2 for the TCR-α locus; and *Bam*HI/pB for the TCR-β locus. Probes and fragment lengths (kb) are indicated on the sides. C57BL/6 × 129/SvJae ES cell DNA served as a wild-type control. Asterisks indicate rearrangement bands. **b**, **c**, Southern blot analyses of different tissues of LN1 (**b**) and LN2 (**c**) mice to show monoclonality. Genomic DNA was digested with *Bam*HI and probed with a mixture of probes pH1 and pB (LN1 tissues), or digested with *Eco*RI and probed with a mixture of probes pA1 and pA2 (LN2 tissues). Organs from C57BL/6 × DBA/2 or 129/SvJae × C57BL/6 mice were used as wild-type controls for LN1 and LN2 mice, respectively. **d–g**, Sequenced joints of the IgH, IgL and TCR-β rearrangements. Single elements are underlined and labelled. Asterisks mark P/N nucleotides not encoded in the germline. WT, wild type; ES, embryonic stem cell.



**Figure 2** Immunological analyses of B- and T-cell-derived mice. **a**, Detection of lymphocytes in LN1 and LN2 mice. PBLs (LN1 mice) or splenocytes (LN2 mice) were analysed by FACS for the expression of markers specific to B cells (B220/IgM) and T cells (CD4/CD8). B6D2F1 (C57BL/6 × DBA/2) mice and 129B6F1 (129/SvJae × C57BL/6) mice served as wild-type controls. The LN1/4n mouse was derived by tetraploid embryo complementation; LN2/Rag-2 was a chimaera generated in a *Rag-2*<sup>-/-</sup> background. **b**, Haplotype and isotype expression of IgH and IgL chains, respectively, in LN1 mice. PBLs of mice derived by tetraploid embryo complementation were stained with haplotype-specific antibodies detecting IgM produced from either the C57BL/6 (IgM<sup>b</sup>) or DBA/2 (IgM<sup>a</sup>) allele (column 1). Splenocytes of LN1/Rag-2 chimaeras were analysed for the expression of Igκ versus Igλ isotypes (columns 2 and 3). **c**, Expression of the initially rearranged variable element in splenocytes derived from LN1/Rag-2 chimaeras compared with a wild-type (WT) control. RNA was analysed by Northern blot analysis using a probe recognizing the rearranged V(H) element. Equal loading was confirmed by rehybridization of the blot with a GAPDH (GD) probe. **d**, Expression of V14 on TCR-β-positive T cells in LN2/Rag-2 complementation chimaeras (column 1) and comparison of cell numbers of αβ T cells versus γδ T cells in the thymus of LN2/Rag-2 chimaeras (columns 2 and 3). Numbers in quadrants represent the percentage of gated lymphocytes.



**Table 2 Generation of mice**

| ES cell line | 4n blastocysts injected | Pups born alive (died) | 2n blastocysts injected | Pups born alive (died) |
|--------------|-------------------------|------------------------|-------------------------|------------------------|
| LN1          | 159                     | 16 (3)                 | 21                      | 14 (0)                 |
| LN2          | 101                     | 1 (1)                  | 105                     | 7 (4)                  |

Generation of entirely ES-cell-derived and chimaeric mice from ES cell lines LN1 and LN2. All 2n blastocysts were deficient for the *Rag-2* gene. Numbers in brackets indicate the number of pups that died neonatally. 2n, diploid; 4n, tetraploid.

genomic DNA isolated from organs of these animals revealed the absence of a germline band for the IgH locus, confirming that these animals were monoclonal and carried the somatic rearrangements of the donor nucleus in all tissues (Fig. 1b). The cloned animals were fertile, gave normal-sized litters and showed no obvious phenotype under normal housing conditions. The ES cells were also injected into diploid blastocysts deficient for recombination-activating gene 2 (*Rag-2*) (Table 2). Because the *Rag-2* mutation blocks lymphocyte maturation, all mature lymphocytes in a chimaera are derived from the donor ES cells<sup>24</sup>.

To assess whether normal numbers of B and T cells were present in mice derived from ES cell line LN1, fluorescence-activated cell sorting (FACS) was performed on peripheral blood lymphocytes (PBLs) using antibodies recognizing surface markers specific to B and T cells. Normal counts of cells positive for B220/IgM (immunoglobulin- $\mu$ ) and the CD4/CD8 antigens were detected in LN1 mice (Fig. 2a, compare columns 1 and 2), consistent with the phenotype of transgenic mice expressing a single immunoglobulin receptor<sup>25</sup>. The positive staining of cells for IgM demonstrates that the donor B cell from which these mice were derived had not yet undergone 'class switching'<sup>17</sup>. Staining of PBLs with haplotype-specific antibodies against IgM revealed that only the 'a' haplotype was expressed on B lymphocytes of LN1-derived mice, indicating that the DBA/2 allele had been chosen for productive IgH rearrangement in the donor B cell (Fig. 2b, column one). Offspring of LN1 mice that carried the rearranged DBA/2 allele over a wild-type allele expressed IgM only from the DBA/2 allele, demonstrating 'allelic exclusion'<sup>17</sup>. Conversely, offspring that carried the nonfunctional IgH allele over a wild-type allele expressed IgM only from the wild-type allele (data not shown). Moreover, all B cells of LN1/*Rag-2* chimaeras expressed  $\kappa$  (Igk) but not  $\lambda$  (Igl) light chains on their surface (Fig. 2b, columns 2 and 3), consistent with Southern blot and PCR analyses. Northern blot analysis showed strong expression of the initial V(H) rearrangement in splenocytes (Fig. 2c). In addition, PCR with reverse transcription (RT-PCR) of splenocyte RNA revealed that ten of ten sequenced complementary DNA clones contained the initial V(H) rearrangement, demonstrating monospecificity of most B cells (data not shown). Similarly, we detected only one of the two rearranged  $\kappa$  alleles (*kk4-Jk1*) by RT-PCR, suggesting that this was the functional allele in the donor nucleus.

The derivation of mice from LN2 ES cells was much less efficient. Of 101 blastocysts injected with the LN2 line, one nonviable pup was delivered by caesarean section, demonstrating that the T-cell-derived ES cell line is also fully pluripotent and can give rise to all somatic cell lineages (Table 2). The karyotype of the LN2 line did not reveal any abnormalities (data not shown), consistent with previous observations that ES cells can suffer subtle genetic or epigenetic alterations that impede postnatal survival<sup>5</sup>. Organs from the single pup were analysed by Southern blot analysis for a TCR- $\alpha$ -specific germline band (Fig. 1c). No germline-specific signal was detected in any organ, indicating that the mouse was completely derived from ES cells. Injection of LN2 ES cells into diploid *Rag-2*<sup>-/-</sup> blastocysts resulted in seven live pups, three of which survived to adulthood (Table 2). Southern blot analysis of genomic tail DNA revealed a contribution of ES cells in these chimaeras of 5–35% (data not shown). *Rag-2* complementation chimaeras had normal-sized thymi and wild-type numbers of total thymocytes, splenocytes and lymph node cells (data not shown). When splenocytes of LN2 mice were analysed by FACS for the presence of mature

lymphocytes, normal numbers of B and T cells were detected (Fig. 2a, columns 3 and 4). However, we observed a significant increase in the ratio of CD4 to CD8 single-positive T cells, with a fourfold decrease in the number of CD8-positive cells and a twofold increase in the number of CD4-positive cells. This result was seen in lymphocytes from peripheral blood, spleen and lymph nodes.

When RNA isolated from splenocytes was analysed for TCR- $\beta$  rearrangements by RT-PCR, we detected a single expressed TCR- $\beta$  rearrangement composed of a V14 element (data not shown). Sequencing of the rearrangement revealed the use of a D $\beta$ 1 and a J $\beta$ 1.5 element, confirming our results with genomic DNA (Fig. 1g). Monospecific expression was further confirmed by FACS analysis of peripheral lymphocytes using a V14-specific antibody (Fig. 2d, column 1). We found that most thymocytes of LN2/*Rag-2* mice expressed an  $\alpha\beta$  TCR (76% versus 11% in the wild type) and no  $\gamma\delta$  TCR (Fig. 2d, columns 2 and 3). Owing to the complexity of the TCR- $\alpha$  locus, we were unable to determine the original rearrangements in ES cell line LN2. However, by using primers that detect expressed rearrangements of all 20 V $\alpha$  families, we obtained amplification products for 17 of them. This result can be explained by 'receptor editing' of the initial rearrangement(s) and is consistent with transgenic studies<sup>23</sup>.

The characteristic genetic rearrangements of mature B and T cells provide an unequivocal demonstration for successful reprogramming of terminally differentiated cells. The two-step approach used to produce the monoclonal mice does not demonstrate that B- and T-cell nuclei can be directly reprogrammed to generate the extra-embryonic lineages. However, previous experiments have indicated that the potency of ES cell nuclei to generate the extra-embryonic as well as the embryonic lineages is high after nuclear transfer<sup>3,4,6</sup>. The frequency of generation of cloned blastocysts from lymphocytes was roughly ten times lower than that of cloned blastocysts from fibroblast or cumulus donor cells<sup>22</sup>. These results suggest that reprogramming a terminally differentiated T- or B-cell nucleus is less efficient than reprogramming nuclei from other donor cell populations. Alternatively, the donor nuclei of various cell types may exhibit different sensitivities to physical damage during the nuclear transfer procedure, affecting the overall cloning efficiencies. Different cloning efficiencies have been observed in certain livestock species using alternative techniques and cell types<sup>26</sup>.

The difficulty of reprogramming differentiated B and T cells is consistent with the notion that many previously cloned animals may actually be derived from the nuclei of rare somatic stem cells present in heterogeneous donor cell populations rather than from differentiated cells as has been assumed<sup>8,9</sup>. Although the proportion of stem cells in a given somatic cell population is low, they may, like ES cells, be more easily reprogrammed<sup>3–6,8,9</sup> and thus selectively serve as the donors for surviving clones. It will be important to stably mark the nuclei of somatic donor cell populations such as fibroblast or cumulus cells to assess the validity of this hypothesis. Furthermore, comparing the cloning potential of donor nuclei from haematopoietic stem cells or neural stem cells<sup>8</sup> with that of nuclei from terminally differentiated cells should be informative. It is possible that the derivation of mice by direct transfer of blastocysts cloned from mature lymphocytes into recipient mothers may be extremely inefficient or even impossible. This is consistent with previous unsuccessful attempts to generate cloned animals from spleen or thymus donor cells<sup>7</sup>. The epigenetic state of the genes crucial for embryonic development in the terminally differentiated

lymphocyte nucleus may be in a conformation that cannot or can only rarely be reprogrammed on transfer into the oocyte. Using an ES cell intermediate may be advantageous because faithful reactivation of embryonic genes may be under fewer time constraints than after direct transfer of the cloned blastocyst into the uterus. Alternatively, the culturing of a cloned blastocyst may select for the few cells within that embryo that have successfully reactivated embryonic genes.

The derivation of mice that carry the genetic alterations of the donor nucleus in all tissues demonstrates the general utility of the nuclear-cloning approach to assess whether genetic or epigenetic mechanisms are involved in cellular differentiation and transformation. For example, nuclear transfer would discern whether genetic rearrangements are involved in the mono-allelic activation of a single olfactory receptor gene during maturation of the olfactory system<sup>27</sup>. □

## Methods

### Nuclear transfer and donor cell preparation

Nuclear transfer was performed as described previously<sup>4,6,7,15</sup>. Briefly, oocytes were flushed from superovulated C57BL/6 × DBA/2 F<sub>1</sub> female mice, enucleated and injected with lymphocyte nuclei using a piezoelectric micromanipulator (Primetech). One to three hours after nuclear transfer, oocytes were activated for 6 h with 10 mM Sr<sup>2+</sup> in Ca<sup>2+</sup>-free medium in the presence of 5 µg ml<sup>-1</sup> of cytochalasin B. Lymphocytes were isolated from lymph nodes of C57BL/6 × DBA/2 F<sub>1</sub> or 129/SvJae × C57BL/6 F<sub>1</sub> male mice using frosted microscope slides (Corning), and cell suspensions were filtered through 50-µm nylon mesh before use.

### Embryo culture and ES cell derivation

Cloned embryos were cultured in MCZB (modified Chatot Ziomek Bavister) medium as described previously<sup>6,7,15</sup>. Blastocysts were briefly treated with Acid Tyrode's solution to remove the zona pellucida and then transferred onto γ-irradiated primary feeder fibroblasts in DMEM (15% FBS, 1,000 U ml<sup>-1</sup> of leukaemia inhibitory factor, LIF). Inner cell masses were dissociated after 3–4 d with trypsin-EDTA and replated onto feeder cells to establish stable ES cell lines. ES cell line LN2 was derived in the presence of the MEK1 kinase inhibitor PD98059 (New England Biolabs) to prevent differentiation of ES cells.

### Generation of mice

Mice were generated by tetraploid embryo complementation and conventional diploid blastocyst injections using a piezoelectric micromanipulator. Tetraploid embryo complementation was performed as described<sup>6</sup>. Diploid blastocysts were derived from 129/SvEv × C57BL/6 F<sub>2</sub> zygotes deficient for the *Rag-2* gene. We injected 8–20 ES cells into blastocysts, which were then transferred into day 2.5 pseudo-pregnant Swiss recipient females. Caesarean sections were performed on day 19.5 and live pups were fostered.

### RNA analysis

Total RNA from organs and cells was extracted with TRI Reagent (Molecular Research Center). We separated 10 µg of RNA on a 1% agarose-formaldehyde gel and blotted them onto Genescreen Plus (NEN) filters, which were then hybridized overnight in Church buffer. The northern probe was generated by random hexamer labelling of a 300-base-pair genomic PCR fragment amplified from the rearranged V(H) element. For RT-PCR, 2 µg of RNA were reverse transcribed using the Superscript Preamplification System (Gibco BRL). To detect the expressed TCR-β rearrangements, a panel of 19 oligonucleotides recognizing each of the published V elements, in combination with nested primers in the constant region 1, was used. Each primer pair amplified a specific product of the expected size using wild-type thymus cDNA as a template. Primer sequences are available on request.

### DNA analysis

Southern blots were hybridized with a 1.3-kilobase (kb) *HindIII*–*XhoI* fragment upstream of DQ52 or a 0.2-kb *PstI*–*HindIII* fragment downstream of JH3 to detect IgH rearrangements, and a 0.7-kb *HindIII* fragment upstream of Jκ1 to detect IgL rearrangements. TCR-α rearrangements were analysed with a 0.4-kb and a 0.5-kb Jα-specific probe, corresponding to probes 9 and 13 (ref. 28). TCR-β rearrangements were detected using a 0.9-kb *Clal*–*EcoRI* probe binding downstream of Jβ2 and corresponding to probe 'D'<sup>29</sup>. Loading of the blot detecting IgL rearrangements was controlled with a 0.6-kb *EcoRI*–*HindIII* fragment downstream of JH4. Sequencing of the rearranged IgH and IgL(κ) loci was performed on ES cell DNA using degenerate primers and PCR conditions as described elsewhere<sup>30</sup>. D-to-J rearrangements at the TCR-β locus were detected by PCR analysis<sup>29</sup>. PCR products were separated on agarose gels and isolated using the QIAGEN gel extraction kit. Purified fragments were then cloned into the TOPO-pCR2.1 vector (Invitrogen) and sequenced using M13 forward and reverse primers.

### FACS analysis

PBLs and splenocytes were treated with ACK lysing buffer (0.15 mM NH<sub>4</sub>Cl, 10 mM KHCO<sub>3</sub>, and 0.1 mM sodium EDTA at pH 7.2) before FACS analysis to remove red blood cells. We stained 1 × 10<sup>6</sup> cells with phycoerythrin (PE)-B220 and fluorescein isothiocyanate (FITC)-IgM antibodies to detect B cells; FITC-CD4 and PE-CD8

antibodies to detect T cells; FITC-IgM<sup>a</sup> and PE-IgM<sup>b</sup> antibodies to determine haplotype-specific IgM expression; FITC-Igκ and FITC-Igλ1/λ2/λ3 antibodies to detect isotype expression; and Cy-Chrome-TCR-β, PE-TCR-γδ, FITC-CD3-ε and FITC-V14 antibodies to detect TCR expression. All antibodies were purchased from Pharmingen. FACS analyses were performed on a Becton Dickinson cell sorter.

Received 10 October 2001; accepted 15 January 2002.

Published online 10 February 2002, DOI 10.1038/nature718.

1. Rideout, W. M. III, Eggan, K. & Jaenisch, R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* **293**, 1093–1098 (2001).
2. Solter, D. Mammalian cloning: advances and limitations. *Nature Rev. Genet.* **1**, 199–207 (2000).
3. Wakayama, T., Rodriguez, I., Perry, A. C., Yanagimachi, R. & Mombaerts, P. Mice cloned from embryonic stem cells. *Proc. Natl Acad. Sci. USA* **96**, 14984–14989 (1999).
4. Rideout, W. M. *et al.* Generation of mice from wild-type and targeted ES cells by nuclear cloning. *Nature Genet.* **24**, 109–110 (2000).
5. Humpherys, D. *et al.* Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95–97 (2001).
6. Eggan, K. *et al.* Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl Acad. Sci. USA* **98**, 6209–6214 (2001).
7. Wakayama, T. & Yanagimachi, R. Mouse cloning with nucleus donor cells of different age and type. *Mol. Reprod. Dev.* **58**, 376–383 (2001).
8. Weissman, I. L. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100**, 157–168 (2000).
9. Liu, L. Cloning efficiency and differentiation. *Nature Biotechnol.* **19**, 406 (2001).
10. Gurdon, J., Laskey, R. & Reeves, O. The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. *J. Embryol. Exp. Morphol.* **34**, 93–112 (1975).
11. DiBerardino, M. Genetic stability and modulation of metazoan nuclei transplanted into eggs and oocytes. *Differentiation* **17**, 17–30 (1980).
12. Briggs, R. & King, T. Changes in the nuclei of differentiating endoderm cells as revealed by nuclear transplantation. *J. Morphol.* **100**, 269–311 (1957).
13. McCreath, K. J. *et al.* Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* **405**, 1066–1069 (2000).
14. Wabl, M. R., Brun, R. B. & Du Pasquier, L. Lymphocytes of the toad *Xenopus laevis* have the gene set for promoting tadpole development. *Science* **190**, 1310–1312 (1975).
15. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
16. Galli, C. Mammalian leukocytes contain all the genetic information necessary for the development of a new individual. *Cloning* **1**, 161–170 (1999).
17. Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **381**, 751–758 (1996).
18. Kisielow, P. & von Boehmer, H. Development and selection of T cells: facts and puzzles. *Adv. Immunol.* **58**, 87–209 (1995).
19. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
20. Kawase, E., Yamazaki, Y., Yagi, T., Yanagimachi, R. & Pedersen, R. A. Mouse embryonic stem (ES) cell lines established from neuronal cell-derived cloned blastocysts. *Genesis* **28**, 156–163 (2000).
21. Munzie, M. J. *et al.* Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. *Curr. Biol.* **10**, 989–992 (2000).
22. Wakayama, T. *et al.* Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. *Science* **292**, 740–743 (2001).
23. Kouskoff, V. & Nemazee, D. Role of receptor editing and revision in shaping the B and T lymphocyte repertoire. *Life Sci.* **69**, 1105–1113 (2001).
24. Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. *RAG-2*-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).
25. Cascalho, M., Ma, A., Lee, S., Masat, L. & Wabl, M. A quasi-monoclonal mouse. *Science* **272**, 1649–1652 (1996).
26. Wells, D. N., Misica, P. M. & Tervit, H. R. Production of cloned calves following nuclear transfer with cultured adult mural granulosa cells. *Biol. Reprod.* **60**, 996–1005 (1999).
27. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
28. Livak, F., Petrie, H. T., Crispe, I. N. & Schatz, D. G. In-frame TCRδ gene rearrangements play a critical role in the αβ/γδ T cell lineage decision. *Immunity* **2**, 617–627 (1995).
29. Whitehurst, C. E., Chattopadhyay, S. & Chen, J. Control of V(D)J recombination accessibility of the Dβ1 gene segment at the TCRβ locus by a germline promoter. *Immunity* **10**, 313–322 (1999).
30. Novobrantseva, T. I. *et al.* Rearrangement and expression of immunoglobulin light chain genes can precede heavy chain expression during normal B cell development in mice. *J. Exp. Med.* **189**, 75–88 (1999).

## Acknowledgements

We thank W. Rideout, K. Eggan, D. Humpherys and Z. Wang for discussions on the project; S. Nguyen, V. Haase, T. Novobrantseva and C. J. Ryu for advice with immunological analyses and for Southern blot probes; J. Dausman and R. Flannery for animal care; and A. Chess, C. Beard and J. Gribnau for critical reading of the manuscript. K.H. was supported by a PhD fellowship from the Boehringer Ingelheim Fonds. Support to R.J. was from the National Cancer Institute.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to R.J. (e-mail: jaenisch@wi.mit.edu).

# MEC-2 regulates *C. elegans* DEG/ENaC channels needed for mechanosensation

Miriam B. Goodman<sup>\*†‡</sup>, Glen G. Ernstrom<sup>\*†</sup>, Dattananda S. Chelur<sup>\*</sup>, Robert O'Hagan<sup>\*</sup>, C. Andrea Yao<sup>\*</sup> & Martin Chalfie<sup>\*</sup>

<sup>\*</sup> Department of Biological Sciences, Columbia University, New York 10027, USA

<sup>†</sup> These authors contributed equally to this work

Touch sensitivity in animals relies on nerve endings in the skin that convert mechanical force into electrical signals. In the nematode *Caenorhabditis elegans*, gentle touch to the body wall is sensed by six mechanosensory neurons<sup>1</sup> that express two amiloride-sensitive Na<sup>+</sup> channel proteins (DEG/ENaC). These proteins, MEC-4 and MEC-10, are required for touch sensation and can mutate to cause neuronal degeneration<sup>2,3</sup>. Here we show that these mutant or 'd' forms of MEC-4 and MEC-10 produce a constitutively active, amiloride-sensitive ionic current when co-expressed in *Xenopus* oocytes, but not on their own. MEC-2, a stomatin-related protein needed for touch sensitivity<sup>4</sup>, increased the activity of mutant channels about 40-fold and allowed currents to be detected with wild-type MEC-4 and MEC-10. Whereas neither the central, stomatin-like domain of MEC-2 nor human stomatin retained the activity of full-length MEC-2, both produced amiloride-sensitive currents with MEC-4d. Our findings indicate that MEC-2 regulates MEC-4/MEC-10 ion channels and raise the possibility that similar ion channels may be formed by stomatin-like proteins and DEG/ENaC proteins that are co-expressed in both vertebrates and invertebrates<sup>5–8</sup>. Some of these channels may mediate mechanosensory responses.

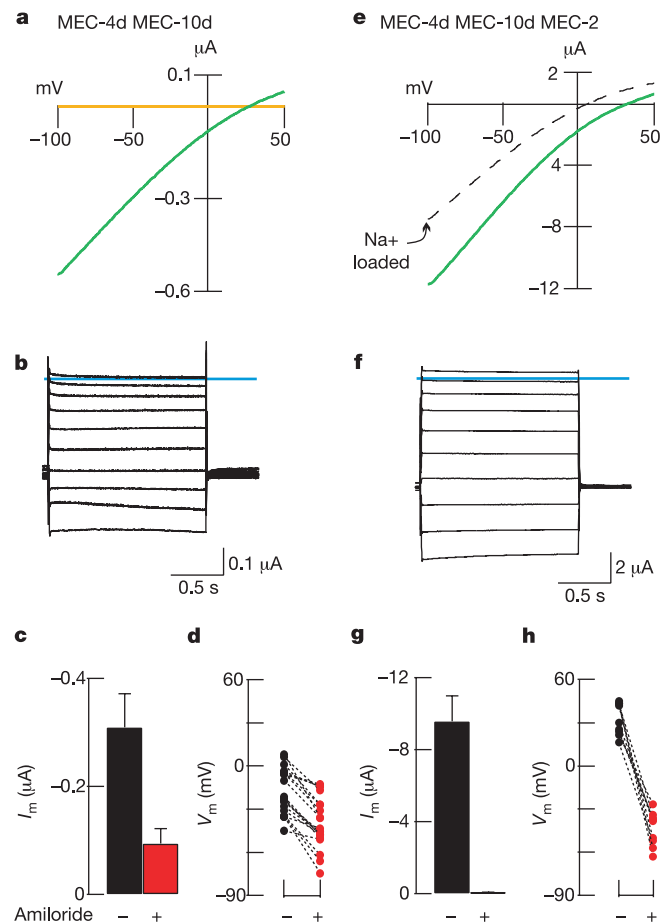
MEC-4 and MEC-10, which are 53% identical, function non-redundantly in mechanosensation<sup>1,9</sup>. As found for related DEG/ENaC channels from worms<sup>10</sup>, flies<sup>11</sup>, and humans<sup>12</sup>, no amiloride-sensitive current was detected in oocytes expressing one or both wild-type proteins (see below). The 'd' forms, by contrast, produced amiloride-sensitive currents only when expressed together (Figs 1a and 2a), even though both proteins were present in the plasma membrane when expressed alone (Fig. 2b). The requirement for both proteins is not due to a change in the amount of surface protein; the ratio of MEC-4d versus MEC-4d/MEC-10d was 0.98 ( $n = 2$ ) and for MEC-10d versus MEC-4d/MEC-10d it was 1.04 ( $n = 2$ ). The current produced by MEC-4d/MEC-10d displayed a mild inward-rectification and lacked any obvious time-dependent component (Fig. 1b). These results suggest that gain-of-function mutations that cause cell death *in vivo* and that activate other DEG/ENaC channels<sup>10–12</sup> also activate MEC-4/MEC-10. Consistent with this idea, MEC-4d/MEC-10d increased the resting membrane potential ( $V_m$ ), and made  $V_m$  sensitive to amiloride (Fig. 1d).

The MEC-4d/MEC-10d current was carried by Na<sup>+</sup> ions, because it reversed polarity at  $15 \pm 3$  mV ( $n = 18$ ) and was essentially eliminated by substituting K<sup>+</sup> for Na<sup>+</sup> in the external saline (not shown). Like  $\alpha\beta\gamma$ ENaC (ref. 13), this current was more permeable to Li<sup>+</sup> than Na<sup>+</sup> ( $P_{Li}/P_{Na} = 3.1 \pm 0.5$ ,  $n = 6$ ). The permeability of the MEC-4d/MEC-10d current differs from that of the *C. elegans* DEG/ENaC protein UNC-105d, which forms channels that are less permeable to lithium ions and more permeable to potassium ions<sup>10</sup>. Similarly, the apparent amiloride inhibition constant ( $K_i$ ), which was  $0.12 \pm 0.03$   $\mu$ M ( $n = 6$ ) at  $-100$  mV, was similar to that of  $\alpha\beta\gamma$ ENaC (ref. 13), but significantly smaller than that of UNC-105d<sup>10</sup>. Thus, the properties of MEC-4d/MEC-10d currents more

closely resemble those of the mammalian  $\alpha\beta\gamma$ ENaC channel than the *C. elegans* UNC-105 channel.

Genetic interactions suggest that MEC-2, which is expressed in all six touch cells<sup>4</sup>, regulates MEC-4/MEC-10 ion channels<sup>3</sup>. We tested for functional interactions by co-expressing MEC-2 with MEC-4d and MEC-10d in *Xenopus* oocytes (Fig. 1e–h). MEC-2, which had no effect on membrane current when expressed alone (Fig. 2c), increased the amplitude of amiloride-sensitive currents about 40-fold but did not affect their voltage- or time-dependence (compare Fig. 1c with Fig. 1g). MEC-2 may regulate the ion permeation pathway, because it reduced the relative permeability for Li<sup>+</sup> ions ( $P_{Li}/P_{Na} = 1.44 \pm 0.07$ ,  $n = 16$ ). Interactions between genes encoding UNC-1 (a stomatin-like protein) and UNC-8 (a DEG/ENaC protein)<sup>14</sup> suggest that this activity is shared by other stomatin-like proteins in *C. elegans*.

The sodium current produced by co-expressing MEC-4d, MEC-10d and MEC-2 drives  $V_m$  towards the expected Nernst potential for Na<sup>+</sup> ions,  $E_{Na}$  ( $V_m = +32 \pm 4$  mV,  $n = 9$ ) in oocytes cultured in the presence of 300  $\mu$ M amiloride. Oocytes cultured without amiloride, by contrast, had resting potentials close to 0 mV ( $V_m = -1.34 \pm 1$  mV,  $n = 30$ ). This observation is reminiscent of the "Na<sup>+</sup> loading" effect described for  $\alpha\beta\gamma$ ENaC channels<sup>13</sup> and is



**Figure 1** MEC-4d, MEC-10d, and MEC-2 produce amiloride-sensitive currents. **a, e**, Voltage-dependence of amiloride difference currents. **a**, MEC-4d/MEC-10d (green), MEC-4d and MEC-10d alone (yellow). **e**, MEC-4d, MEC-10d and MEC-2 in oocytes cultured with (solid green line) and without amiloride (dotted black line). **b, f**, Time-dependence of currents evoked by voltage pulses between  $-100$  and  $+35$  mV (15-mV increments). Zero current (blue). **c, d, g, h**, Membrane current (at  $-85$  mV) and voltage in presence (red) and absence (black) of 300  $\mu$ M amiloride. **c, d**, MEC-4d/MEC-10d ( $n = 23$ ). **g, h**, MEC-4d/MEC-10d and MEC-2 ( $n = 9$ ).  $V_{hold} = -60$  mV; cells cultured with 300  $\mu$ M amiloride, except as indicated.

<sup>‡</sup> Present address: Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305-5345, USA.



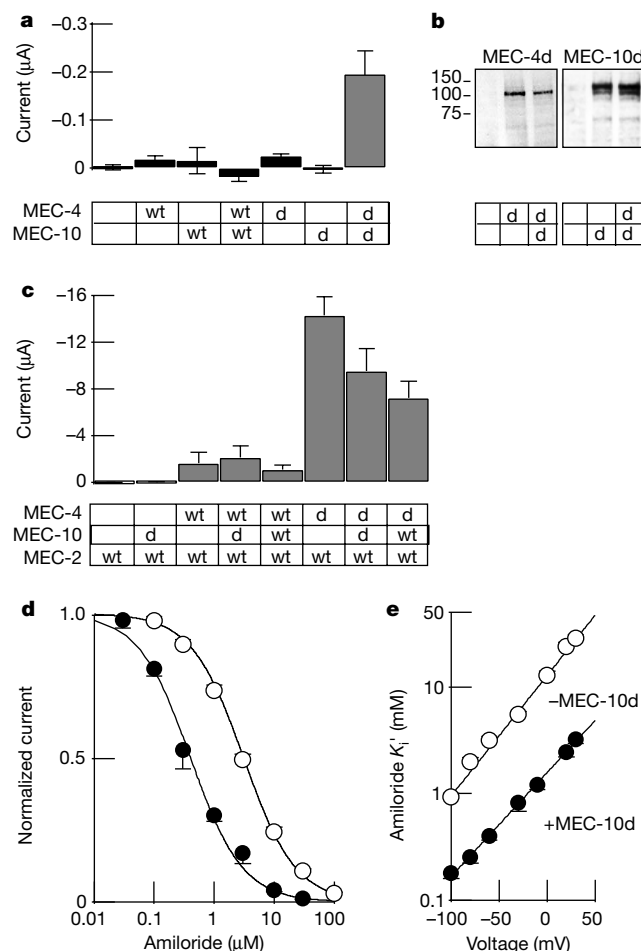
reflected in a shift in the reversal potential of the amiloride-sensitive current (Fig. 1b; control  $E_{rev} = 32 \pm 1$  mV,  $n = 9$  versus  $Na^+$  loaded  $E_{rev} = 4 \pm 2$  mV,  $n = 27$ ). Thus, culturing oocytes without amiloride drives  $E_{Na}$  close to 0 mV.

The amplification provided by MEC-2 allowed us to detect amiloride-sensitive currents in some oocytes (13 of 33 cells from 3 of 6 frogs) expressing wild-type MEC-4 and MEC-10 (Fig. 2c), which produced no amiloride-sensitive current when expressed alone or together (Fig. 2a). This current was observed in the absence of any explicit mechanical stimulation. Current could not be induced in oocytes lacking a constitutive, amiloride-sensitive current by superfusion with salines having either reduced pH (5.2,  $n = 3$ ) or osmolarity (115 mOsm,  $n = 16$ ). This result implies that wild-type channels with MEC-2 may be partially open at rest (that is, open probability,  $P_o > 0$ ). A similar situation may occur *in vivo*. Even a tiny resting  $Na^+$  current would depolarize touch cells, because they exhibit an unusually high input resistance ( $>2$  G $\Omega$ ; M.B.G., S. Lockery, D. Lenzi and M.C., unpublished work). In this case, a change in mechanical force could hyperpolarize touch cells by closing channels and/or depolarize them further by opening channels. Alternatively, other proteins required for touch cell function *in vivo* might decrease resting  $P_o$ . In particular, compo-

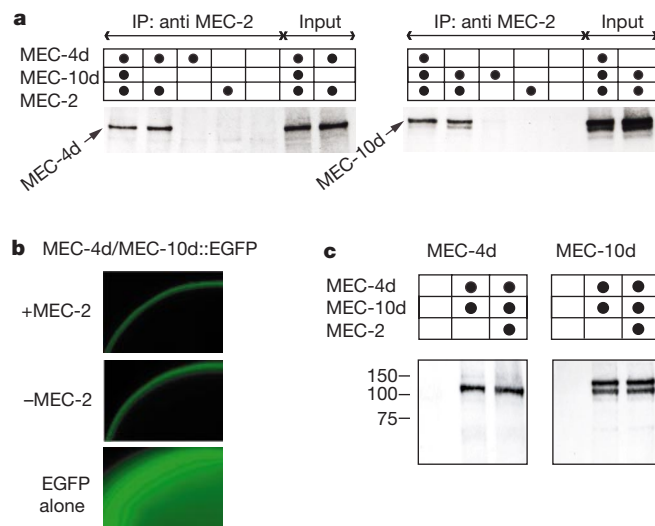
nents of the touch cell extracellular matrix and/or cytoskeleton, which are required for touch sensitivity *in vivo*<sup>1,9</sup>, might prevent the channel from opening at rest. The response to mechanical force would remove inhibition generated by interaction with specialized structures and allow channels to assume their resting  $P_o$ , resulting in depolarization. In this scenario, the channel would not be directly mechanically gated, but would be mechanically sensitive by virtue of its interaction with other proteins *in vivo*.

MEC-4 and MEC-10 exhibit functional differences when co-expressed with MEC-2 in *Xenopus* oocytes. Specifically, introducing the 'd' mutation into MEC-4, but not MEC-10, significantly increased current amplitude (Fig. 2c), a difference that may account for the comparatively weak degeneration phenotype observed with *med-10d* (ref. 3). We also found that MEC-4, but not MEC-10, was both necessary and sufficient to produce amiloride-sensitive currents in the presence of MEC-2 (Fig. 2c). This result agrees with genetic studies showing that *mec-4* is required for neuronal degeneration caused by *mec-10d*, but not *vice versa*<sup>3</sup>, but raises the question of whether or not each protein forms distinct channels or a single heteromeric channel.

We measured amiloride dose-response curves to answer this question and found that adding MEC-10d reduced  $K'_i$  without introducing a second class of binding sites (Fig. 2d). Scatchard plots (not shown) were also consistent with the existence of a single class of binding sites in the presence and absence of MEC-10d. These observations indicate that MEC-4d and MEC-10d form a heteromeric channel. We believe that a single amiloride molecule binds to each channel and inhibits current by lodging in the ion pore formed by MEC-4d and MEC-10d, as initially proposed for native ENaC channels<sup>15</sup>. Consistent with this idea, amiloride blockade was steeply voltage-dependent (Fig. 2e). Analysis of the relation between  $K'_i$  and voltage revealed that amiloride penetrated at least 50% of the membrane electric field (Fig. 2e,  $\delta = 0.54$ – $0.62$ ). These values are similar to those measured for channels formed by UNC-105d ( $\delta = 0.65$ – $0.68$ ) (ref. 10), but greater than those of the  $\alpha\beta\gamma$ ENaC channels ( $\delta = 0.15$ ) (ref. 16). Although the molecular basis for this difference is unknown, it could reflect differences in the accessibility



**Figure 2** Functional interactions of MEC-4, MEC-10 and MEC-2. **a**, Amiloride-sensitive current amplitude (measured at  $-85$  mV in 3–88 cells cultured with and without amiloride). **a**, Wild-type (wt) and 'd' (d) forms. **c**, With MEC-2. **b**, Surface expression of Myc::MEC-4d and MEC-10d::EGFP. **d**, Amiloride dose-response curves (at  $-60$  mV) for MEC-4d, MEC-10d and MEC-2 ( $K'_i = 0.40$   $\mu$ M,  $n = 21$ , filled circles) and MEC-4d and MEC-2 ( $K'_i = 3.2$   $\mu$ M,  $n = 16$ , open circles). **e**, Voltage-dependence of amiloride blockade. The smooth line is a fit using a Woodhull model<sup>30</sup> ( $\delta = 0.52$  and  $0.63$  in the presence and absence of MEC-10d).



**Figure 3** MEC-2 interacts with MEC-4d and MEC-10d without altering surface expression. **a**, Co-immunoprecipitation of Myc::MEC-4d and MEC-10d::EGFP fusion proteins by antibodies against MEC-2. Five and one oocyte equivalent(s) were loaded in the IP and input lanes, respectively. **b**, Confocal images of live oocytes expressing MEC-4d and MEC-10d::EGFP in the presence (top) and absence (middle) of MEC-2. EGFP (enhanced green fluorescent protein) fluorescence is diffuse (bottom). **c**, Effect of MEC-2 on surface expression of MEC-4d (left) and MEC-10d (right). Each lane represents surface protein from 30–45 oocyte equivalents.

of the amiloride binding site and/or its location with respect to the membrane electric field.

MEC-2 appears to form an ion channel complex with MEC-4 and MEC-10, a result suggested by genetic interactions<sup>3,17</sup>, because antibodies against MEC-2 immunoprecipitated MEC-4d and MEC-10d (Fig. 3a). The interaction is pair-wise: MEC-2 immunoprecipitated MEC-4d in the absence of MEC-10d and *vice versa*. It is also specific, because MEC-2 failed to immunoprecipitate the endogenous membrane protein,  $\beta$ -integrin<sup>18</sup> (not shown). To test whether the increased current size was produced by an increase in the number of MEC-4d/MEC-10d channels that reach the plasma membrane, we tagged MEC-4d and MEC-10d. A MEC-10d::EGFP fusion protein was visible near the plasma membrane of live oocytes (Fig. 3b) and produced amiloride-sensitive currents when co-expressed with MEC-4d and MEC-2 (see Methods). MEC-10d::EGFP localization was not obviously affected by omitting MEC-2. MEC-2 also did not affect the amount of either MEC-4d or MEC-10d available for biotinylation at the surface (Fig. 3c); the ratio of MEC-4d with and without MEC-2 was  $1.2 \pm 0.4$  ( $n = 5$ ) and for MEC-10d it was  $0.9 \pm 0.3$  ( $n = 3$ ). MEC-2 is, therefore, unlikely to increase channel number and probably acts by regulating single channel conductance, open probability, mean open time, or all three parameters.

The central domain of MEC-2 (amino acids 114–363) is 64% identical to stomatin, a human protein implicated in the regulation of ion flux in red blood cells<sup>19</sup>. Fifty-four alleles of *mec-2* were identified in genetic screens for touch-insensitive mutants<sup>1,9</sup>. More than half of these are missense mutations that map to this central, stomatin-like domain (ref. 4 and G. Caldwell and M.C., unpublished work), implying that this domain is especially important for the function of MEC-2. To determine whether stomatin and MEC-2(114–363) function similarly, we co-expressed them with the 'd' form of MEC-4. MEC-2(114–363) produced comparatively small amiloride-sensitive currents (Fig. 4b), indicating that the central domain retains the ability to generate amiloride-sensitive

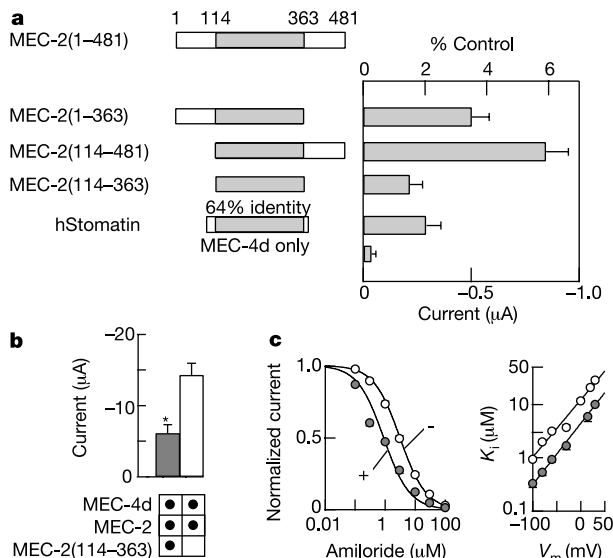
currents with MEC-4d. Stomatin, which had no effect by itself (not shown,  $n = 13$ ), produced amiloride-sensitive currents of a similar size when co-expressed with MEC-4d (Fig. 4a). These findings establish that MEC-2(114–363) and stomatin share the ability to regulate a DEG/ENaC channel and provide the first direct support for the hypothesis that stomatin-like proteins regulate ion channels.

The stomatin-like domain of MEC-2(114–363) reduces current amplitude in a dominant-negative fashion when co-expressed with full-length MEC-2 (Fig. 4b). Human stomatin also produced a strong dominant-negative effect (not shown), reinforcing the functional similarity between the two proteins. Such interference is consistent with the notion that MEC-2 forms multimers via the conserved central domain. This hypothesis is also supported by interallelic complementation at *mec-2* (refs 1, 20) and by physical interactions between stomatin monomers<sup>21</sup>. MEC-2(114–363) also reduced amiloride  $K_i$ , without introducing an additional class of binding sites or changing the voltage-dependence of blockade (Fig. 4c), a finding which suggests that although MEC-2 may regulate access to the amiloride binding site or contribute to its formation, it does not regulate the position of the binding site within the electrical field.

Both unique N-terminal and C-terminal regions of MEC-2, which are believed to be cytoplasmic<sup>3</sup>, are needed for full activity of the protein in *Xenopus* oocytes. Whereas the addition of either terminal domain increased current amplitude to some extent, neither MEC-2(1–363) nor MEC-2(114–481) produced currents as large as those observed with the full-length protein (Fig. 4b). This reduction did not reflect a reduction in protein levels, because both truncated proteins were produced at approximately the same level as the full-length protein (as judged by western blot analysis, not shown). Thus, all three domains contribute to the function of MEC-2.

We propose that the highly conserved, stomatin-like domain of MEC-2 provides an essential structural scaffold for interaction with DEG/ENaC proteins, with the lipids surrounding the channel, or both. Evidence for lipid association comes from the observation that stomatin is palmitoylated *in vivo*<sup>22</sup> and associated with lipid rafts<sup>22,23</sup>, sphingolipid- and cholesterol-rich microdomains in the plasma membrane. The predominant site of palmitoylation in stomatin<sup>22</sup> is conserved in MEC-2 and such a covalent modification, if present, would anchor MEC-2 to the inner leaflet of the plasma membrane. Although MEC-2(114–363) acts in a dominant-negative fashion, the majority of the ability of MEC-2 to regulate ion channel function is explained by the action of the unique N and C termini. The central stomatin-like domain may, therefore, bring these unique domains into close proximity to MEC-4 and MEC-10.

The reconstitution of channel activity in *Xenopus* oocytes establishes the biochemical function of MEC-4, MEC-10 and MEC-2 and is a first step toward understanding the function of other proteins implicated in *C. elegans* mechanosensation. The physical and functional interactions detailed here may provide insight into the function of similar proteins in vertebrates, as well as clues to their role in mechanosensation. A role for DEG/ENaC proteins in vertebrate mechanosensation comes from recent studies showing that BNC1 $\alpha$  (also known as ASIC2a and BNC1) is present in the somata and terminals of dorsal root ganglion (DRG) neurons that innervate mammalian skin<sup>24</sup> and is needed for normal sensory responses in one class of sensory nerve fibres<sup>25</sup>. Stomatin may regulate the channel containing BNC1, as it is expressed in all DRG neurons<sup>6</sup>. Stomatin is also co-expressed with  $\alpha\beta\gamma$ ENaC channels in trigeminal sensory neurons that sense whisker deflections in rats<sup>5</sup> and may regulate these channels. We have shown that human stomatin can partially substitute for MEC-2 in the generation of amiloride-sensitive currents when co-expressed with MEC-4d in *Xenopus* oocytes, a result which supports the idea that stomatin-like proteins share the common function of regulating DEG/ENaC ion channels. Such interactions would expand the



**Figure 4** Three domains are needed for full MEC-2 function. **a**, Activity of truncated MEC-2 and human stomatin. Amiloride-sensitive current (at  $-85$  mV in 8–32 cells) produced by co-expression with MEC-4d (bottom axis) and compared to full-length MEC-2 (top axis, % control). Dominant-negative effect of MEC-2(114–363) on **b**, current amplitude (asterisk,  $P < 0.01$ ) and **c**, amiloride sensitivity. Normalized dose-response curves were obtained at  $-60$  mV (left panel,  $n = 4–16$ );  $K_i = 0.93$  and  $3.2 \mu$ M with (filled) and without (open) MEC-2(114–363), respectively. Voltage-dependence of  $K_i$  (right panel) with (filled circles,  $\delta = 0.68$ ) and without (open circles,  $\delta = 0.63$ ) MEC-2(114–363).

combinatorial possibilities for channel activity beyond that previously imagined for DEG/ENaC proteins alone. □

## Methods

### Expression constructs

Wild-type complementary DNAs encoding full-length MEC-2 (ref. 5), truncated MEC-2 proteins, MEC-4 and MEC-10 were subcloned into pGEM-HE or pSGEM with a Kozak sequence upstream of the initial codon. Plasmids encoding wild-type MEC-4 (TU#667) and MEC-10 (TU#668) were mutated *in vitro* to give plasmids encoding MEC-4d (TU#655) and MEC-10d (TU#656). These mutations reproduce the *mec-4(e1611)* (ref. 2) mutation and correspond to A713T and A673T in MEC-4 and MEC-10, respectively. Coding sequences were verified by automated DNA sequencing.

### Bacterial strain for degnerin plasmids

Plasmids containing full-length degnerin cDNAs are toxic to standard *E. coli* strains<sup>3,26</sup>; transformants either form tiny colonies or carry mutant plasmids. We generated an *E. coli* strain, SMC4, that showed normal growth with and stable propagation of *mec-4* and *mec-10* plasmids by randomly mutating *E. coli* NM554 with the mini-Tn10cma transposon<sup>27</sup>, transforming with a *mec-10* plasmid, and screening for normal growth. We tested for stable propagation by showing that the plasmid caused NM554 and XL2blue to give tiny colonies, curing the strain of the plasmid, and testing for growth of a *mec-4* plasmid. TU#667, TU#668, TU#655 and TU#656 and their derivatives were propagated in SMC4.

### Oocyte expression and electrophysiology

Capped RNAs were synthesized (T7 mMESSAGE mMACHINE kit, Ambion), purified, and quantified spectroscopically. *Xenopus laevis* oocytes were collected and injected with 10 ng of each cRNA, except for oocytes co-expressing only MEC-4d and MEC-2 (114–363), which were injected with 10 ng of the former and 20 ng of the latter. Oocytes were maintained in L-15 oocyte medium containing 100 µg ml<sup>-1</sup> gentamicin (Cell & Molecular Technologies) at 16–18 °C. Where indicated, 300 µM amiloride was added to the culture medium.

Membrane potential and current were measured 4–10 days after cRNA injection using a two-electrode voltage clamp (Warner OC-725C) at 22–25 °C. Electrodes (0.3–2 MΩ) were filled with 3 M KCl and oocytes were superfused with saline containing (in mM): Na-glucuronate (100), KCl (2), CaCl<sub>2</sub> (1), MgCl<sub>2</sub> (2), NaHEPES (10), pH 7.2. For low pH experiments, HEPES was replaced by MES. For hypo-osmotic experiments, saline was diluted to 100–110 mOsm. Current was similar in hypo-osmotic saline supplemented with sucrose. Ion selectivity experiments used salines containing (in mM): X<sup>-</sup>-glucuronate (100), CaCl<sub>2</sub> (1), MgCl<sub>2</sub> (2) and X<sup>-</sup>HEPES (10), pH 7.2, where X<sup>-</sup> was Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Cs<sup>+</sup> or N-methyl-D-glucamine (NMG). Amiloride-difference currents were used to determine ion selectivity of MEC-4d/MEC-10d channels. All chemicals were obtained from Sigma.

Analogue signals were filtered at 200 Hz and sampled at 1–2 kHz (ITC-16, Instrutech); a 60-Hz notch filter was used to minimize line noise. Average values are reported as mean ± s.e.m. Curves were fit by a nonlinear least-squares method (IgorPro 4.01, Wavemetrics); the standard deviation measured at each point provided the weighting function. For dose-response relations, current was normalized to the total amiloride-sensitive current (measured as the difference in control and 300 µM amiloride salines). Relative permeabilities were calculated from the difference in reversal potential measured in solutions containing Na<sup>+</sup> and each test ion using the Goldman–Hodgkin–Katz equation<sup>28</sup>.

### Surface expression, co-immunoprecipitation and western blotting

Surface expression was assayed according to ref. 29. Treatment groups were normalized to 30–40 oocyte equivalents per lane and subject to SDS-PAGE followed by western blotting. The C terminus of MEC-10d was fused to EGFP (Clontech). When co-expressed with MEC-4d and MEC-2, this fusion protein generated amiloride sensitive currents [*I*<sub>amil</sub> (–85 mV) = –5.3 ± 1 µA, *n* = 6] with a slightly elevated *K*<sub>i</sub> for amiloride [*K*<sub>i</sub> (–60 mV) = 1.0 ± 0.2 µM, *n* = 7]. Cytoplasmic EGFP was detected in the supernatant (internal) fraction, but not in the streptavidin precipitate (not shown). The N terminus of MEC-4d was fused to Myc; Myc:MEC-4d was functional when co-expressed with MEC-10d and MEC-2 [*I*<sub>amil</sub> (–85 mV) = –9.8 ± 3 µA, *n* = 3]. MEC-10d:EGFP and Myc:MEC-4d were detected either with HRP-conjugated antibodies against the epitope tags (Santa Cruz Biotechnology) or with primary antibodies against the epitope tags (Zymed) and HRP-conjugated secondary antibodies. HRP was detected using chemiluminescence (ECL and ECLplus, Amersham Pharmacia Biotech). Band density was measured from digitized films using NIH Image; intensity was corrected *post hoc* for variation in oocyte equivalents loaded.

Ion channel complexes were immunoprecipitated from oocyte homogenates with rabbit polyclonal antibodies raised against purified, bacterial MEC-2 (145–481) (S. Zhang and M.C., unpublished work). Homogenates were prepared 5–6 days after cRNA injection using 10 µl of lysis buffer (20 mM Tris-HCl, pH 7.6, 100 mM NaCl, 2% NP-40) per oocyte. Yolk platelets were removed by low-speed centrifugation and the supernatant diluted with lysis buffer to a final concentration of 2–10 oocytes ml<sup>-1</sup>.

Immunocomplexes were precipitated by Protein A/G PLUS conjugated to agarose (Santa Cruz Biotechnology), washed three times in lysis buffer, and analysed by SDS-PAGE. Four to five oocyte equivalents were loaded per 'IP' lane; one oocyte equivalent was loaded per input lane. Western blotting was essentially as described above. We confirmed the specificity of the interaction in two ways: (1) using anti-Myc and anti-EGFP antibodies conjugated to agarose to immunoprecipitate MEC-2 from the same sample (not shown); and (2) probing immuno-complexes for the presence of β-integrin with a monoclonal

antibody (8C8, Developmental Studies Hybridoma Bank, Univ. Iowa), an unrelated, *Xenopus* oocyte membrane protein<sup>18</sup>.

Received 20 September; accepted 21 December 2001.

- Chalfie, M. & Sulston, J. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Dev. Biol.* **82**, 358–370 (1981).
- Driscoll, M. & Chalfie, M. The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. *Nature* **349**, 588–593 (1991).
- Huang, M. & Chalfie, M. Gene interactions affecting mechanosensory transduction in *Caenorhabditis elegans*. *Nature* **367**, 467–470 (1994).
- Huang, M., Gu, G., Ferguson, E. L. & Chalfie, M. A stomatin-like protein necessary for mechanosensation in *C. elegans*. *Nature* **378**, 292–295 (1995).
- Fricke, B. *et al.* Epithelial Na<sup>+</sup> channels and stomatin are expressed in rat trigeminal mechanosensory neurons. *Cell Tissue Res.* **299**, 327–334 (2000).
- Mannsfeldt, A. G., Carroll, P., Stucky, C. L. & Lewin, G. R. Stomatin, a MEC-2 like protein, is expressed by mammalian sensory neurons. *Mol. Cell Neurosci.* **13**, 391–404 (1999).
- Tavernarakis, N., Shreffler, W., Wang, S. & Driscoll, M. unc-8, a DEG/ENaC family member, encodes a subunit of a candidate mechanically gated channel that modulates *C. elegans* locomotion. *Neuron* **18**, 107–119 (1997).
- Sedensky, M., Siefker, J. M. & Morgan, P. G. Model organisms: new insights into ion channel and transporter function. Stomatin homologues interact in *Caenorhabditis elegans*. *Am. J. Physiol. Cell Physiol.* **280**, C1340–C1348 (2001).
- Chalfie, M. & Au, M. Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. *Science* **243**, 1027–1033 (1989).
- García-Añoveros, J., García, J. A., Liu, J. D. & Corey, D. P. The nematode degnerin UNC-105 forms ion channels that are activated by degeneration- or hypercontraction-causing mutations. *Neurons* **20**, 1231–1241 (1998).
- Adams, C. M. *et al.* Ripped pocket and pickpocket, novel *Drosophila* DEG/ENaC subunits expressed in early development and in mechanosensory neurons. *J. Cell Biol.* **140**, 143–152 (1998).
- Waldmann, R., Champigny, G., Voilley, N., Lauritzen, I. & Lazdunski, M. The mammalian degnerin MDEG, an amiloride-sensitive cation channel activated by mutations causing neurodegeneration in *Caenorhabditis elegans*. *J. Biol. Chem.* **271**, 10433–10436 (1996).
- Canessa, C. M. *et al.* Amiloride-sensitive epithelial Na<sup>+</sup> channel is made of three homologous subunits. *Nature* **367**, 463–467 (1994).
- Rajaram, S., Spangler, T. L., Sedensky, M. M. & Morgan, P. G. A stomatin and a degnerin interact to control anesthetic sensitivity in *Caenorhabditis elegans*. *Genetics* **153**, 1673–1682 (1999).
- Palmer, L. G. Interactions of amiloride and other blocking cations with the apical Na channel in the toad urinary bladder. *J. Membr. Biol.* **87**, 191–199 (1985).
- McNicholas, C. M. & Canessa, C. M. Diversity of channels generated by different combinations of epithelial sodium channel subunits. *J. Gen. Physiol.* **109**, 681–692 (1997).
- Gu, G., Caldwell, G. A. & Chalfie, M. Genetic interactions affecting touch sensitivity in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 6577–6582 (1996).
- Muller, A. H., Gawantka, V., Ding, X. & Hausen, P. Maturation induced internalization of beta 1-integrin by *Xenopus* oocytes and formation of the maternal integrin pool. *Mech. Dev.* **42**, 77–88 (1993).
- Lande, W. M., Thiemann, P. V. & Mentzer, W. C. Jr Missing band 7 membrane protein in two patients with high Na, low K erythrocytes. *J. Clin. Invest.* **70**, 1273–1280 (1982).
- Huang, M. Mechanosensory genes in *Caenorhabditis elegans*. PhD thesis, Columbia Univ. (1995).
- Snyers, L., Umlauf, E. & Prohaska, R. Oligomeric nature of the integral membrane protein stomatin. *J. Biol. Chem.* **273**, 17221–17226 (1998).
- Snyers, L., Umlauf, E. & Prohaska, R. Cysteine 29 is the major palmitoylation site on stomatin. *FEBS Lett.* **449**, 101–104 (1999).
- Salzer, U. & Prohaska, R. Stomatin, flotillin-1, and flotillin-2 are major integral proteins of erythrocyte lipid rafts. *Blood* **97**, 1141–1143 (2001).
- García-Añoveros, J., Samad, T. A., Woolf, C. J. & Corey, D. P. Transport and localization of the DEG/ENaC ion channel BNaCl α to peripheral mechanosensory terminals of dorsal root ganglia neurons. *J. Neurosci.* **21**, 2678–2686 (2001).
- Price, M. P. *et al.* The mammalian sodium channel BNC1 is required for normal touch sensation. *Nature* **407**, 1007–1011 (2000).
- Lai, C. C., Hong, K., Kinnell, M., Chalfie, M. & Driscoll, M. Sequence and transmembrane topology of MEC-4, an ion channel subunit required for mechanotransduction in *Caenorhabditis elegans*. *J. Cell Biol.* **133**, 1071–1081 (1996).
- Kleckner, N., Bender, J. & Gottesman, S. Uses of transposons with emphasis on Tn10. *Methods Enzymol.* **204**, 139–180 (1991).
- Hille, B. *Ion Channels of Excitable Membranes*. (Sinauer, Sunderland, Massachusetts, 2001).
- Chillaron, J. *et al.* An intracellular trafficking defect in type I cystinuria rBAT mutants M467T and M467K. *J. Biol. Chem.* **272**, 9543–9549 (1997).
- Woodhull, A. M. Ionic blockage of sodium channels in nerve. *J. Gen. Physiol.* **61**, 687–708 (1973).

### Acknowledgements

We thank S. Zhang for polyclonal anti-MEC-2 (145–481); M. Driscoll for wild-type degnerin cDNAs; G. Stewart for hStomatin cDNA; S. Hollman for pSGEM (a derivative of pGEM-HE); the Developmental Studies Hybridoma Bank for monoclonal antibodies against *Xenopus* β-integrin; J. Art, S. Firestein and J. Yang for the loan of equipment. We also thank L. Chen for technical assistance and T. Lu and J. Yang for assistance with preliminary experiments. This work was supported by a research grant from the National Institute of General Medical Sciences, NIH (M.C.), a Human Frontiers Science Program postdoctoral fellowship (D.S.C.), and an NRSA postdoctoral fellowship from the National Institute of Deafness and Other Communication Disorders, NIH (M.B.G.).

### Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C. (e-mail: mc21@columbia.edu).



# Characterization of a common precursor population for dendritic cells

Gloria Martínez del Hoyo, Pilar Martín, Héctor Hernández Vargas, Sara Ruiz, Cristina Fernández Arias & Carlos Ardavin

Department of Cell Biology, Faculty of Biology, Complutense University, 28040 Madrid, Spain

Dendritic cells (DCs) are essential for the establishment of immune responses against pathogens and tumour cells, and thus have great potential as tools for vaccination and cancer immunotherapy trials. Experimental evidence has led to a dual DC differentiation model, which involves the existence of both myeloid- and lymphoid-derived DCs<sup>1</sup>. But this concept has been challenged by recent reports demonstrating that both CD8<sup>−</sup> and CD8<sup>+</sup> DCs, considered in mice as archetypes of myeloid and lymphoid DCs respectively, can be generated from either lymphoid<sup>2–4</sup> or myeloid progenitors<sup>3,4</sup>. The issue of DC physiological derivation therefore remains an open question. Here we report the characterization of a DC-committed precursor population, which has the capacity to generate all the DC subpopulations present in mouse lymphoid organs—including CD8<sup>−</sup> and CD8<sup>+</sup> DCs, as well as the B220<sup>+</sup> DC subset—but which is devoid of myeloid or lymphoid differentiation potential. These data support an alternative model of DC development, in which there is an independent, common DC differentiation pathway.

Current controversy about the physiological derivation of DCs results largely from the lack of information about DC progenitors. To obtain insights into the developmental origin of DCs, we searched for a murine DC precursor in the blood cell fraction expressing the DC-specific integrin CD11c. Blood CD11c<sup>+</sup> cells can be subdivided into MHC class II (MHC-II)<sup>+</sup> and MHC-II<sup>−</sup> subsets (Fig. 1A, a). Phenotypic analyses (not shown) revealed that CD11c<sup>+</sup> MHC-II<sup>+</sup> cells corresponded to CD4<sup>−</sup> CD8<sup>−</sup> HSA<sup>+</sup> F4/80<sup>+</sup> CD62L<sup>−</sup> blood DCs, similar to splenic CD8<sup>−</sup> DCs (ref. 5), and allowed us to design the method used here to purify the MHC-II<sup>−</sup> subset (Fig. 1A, a–c). CD11c<sup>+</sup> MHC-II<sup>−</sup> cells displayed ultrastructural characteristics of undifferentiated cells (Fig. 1B), and were B220<sup>+</sup>, CD19<sup>−</sup>, CD43<sup>+</sup>, CD11b<sup>+</sup> and 50% were Thy-1<sup>+</sup> (Fig. 1A, b). They expressed Fcγ, CD44 and the homing receptor CD62L, and were negative for the cytokine receptors c-kit, IL-7Rα and IL-3Rα, the B-cell/DC-related molecules CD40, CD86 and DEC-205, the macrophage marker F4/80 and the granulocyte antigen Gr-1. The phenotype of CD11c<sup>+</sup> MHC-II<sup>−</sup> cells, representing around 5% (5.4 ± 0.8, *n* = 4) of blood mononuclear cells, corresponded to that of a progenitor population, as they were negative for lineage markers and expressed early precursor markers, though this phenotypic profile does not allow us to correlate them to lymphoid or myeloid precursors. Importantly, CD11c<sup>+</sup> MHC-II<sup>−</sup> cells neither corresponded with monocytes nor included monocytes, because the last express F4/80 and Gr-1, but are negative for CD11c and CD43. (Comparative phenotype profile of DC-precursors versus monocytes: DC-precursors are CD11c<sup>+</sup> B220<sup>+</sup> CD43<sup>+</sup> HSA<sup>−</sup> Gr-1<sup>−</sup> F4/80<sup>−</sup> whereas monocytes are CD11c<sup>−</sup> B220<sup>−</sup> CD43<sup>−</sup> HSA<sup>+</sup> Gr-1<sup>+</sup> F4/80<sup>+</sup>; C.A., unpublished data). Moreover, DCs were not generated from monocytes after intravenous transfer into irradiated mice (C.A., unpublished data).

We then determined the reconstitution potential of CD11c<sup>+</sup> MHC-II<sup>−</sup> cells transferred into γ-irradiated recipients using an Ly5.1/Ly5.2 system to trace donor-type Ly5.2<sup>+</sup> cells. B220-depleted splenic DC-enriched fractions were analysed for donor-type DCs. Both Ly5.2<sup>+</sup> CD8<sup>−</sup> and CD8<sup>+</sup> DCs were detected after 7 days

(Fig. 2a); by day 14 the number of donor-type DCs was similar to that of DCs in control spleens<sup>5</sup>, indicating that CD11c<sup>+</sup> MHC-II<sup>−</sup> cells had the potential to fully reconstitute the CD8<sup>−</sup> and CD8<sup>+</sup> DC populations. Donor-type DCs displayed similar phenotypic characteristics to control splenic DCs, and their CD8<sup>−</sup>:CD8<sup>+</sup> ratio was 65:35, 50:50 and 35:65 at days 7, 10 and 14 respectively. Because the CD8<sup>−</sup>:CD8<sup>+</sup> ratio of splenic DCs reconstituted from bone marrow precursors was 70:30 after 3 weeks, as in the spleen of control mice<sup>2</sup>, and as CD8<sup>+</sup> DCs have been claimed to derive from the CD8<sup>−</sup> DC subset<sup>6</sup>, our data on DC derivation from CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors suggest the extinction of the CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors during the third week after transfer. This is in agreement with the low DC reconstitution observed 21 days after transfer (not shown). The extinction time of CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors after transfer and their DC reconstitution capacity is in agreement with those described for bone marrow myeloid and lymphoid progenitors (3–4 weeks; refs 3 and 4) and for thymic CD4<sup>low</sup> lymphoid precursors (3 weeks; ref. 7). Analysis of the absolute number of Ly5.2<sup>+</sup> DCs generated from CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors (Fig. 2b) suggests that DC differentiation from this population involves a rapid cell expansion concomitant with the overall splenic reconstitution. (Total spleen cells obtained after reconstitution of irradiated mice with 4 × 10<sup>4</sup> bone marrow cells were as follows: 7 days, 15 × 10<sup>6</sup>; 10 days, 50 × 10<sup>6</sup>; 14 days, 130 × 10<sup>6</sup> cells. Data are from an experiment representative of four with similar results.)

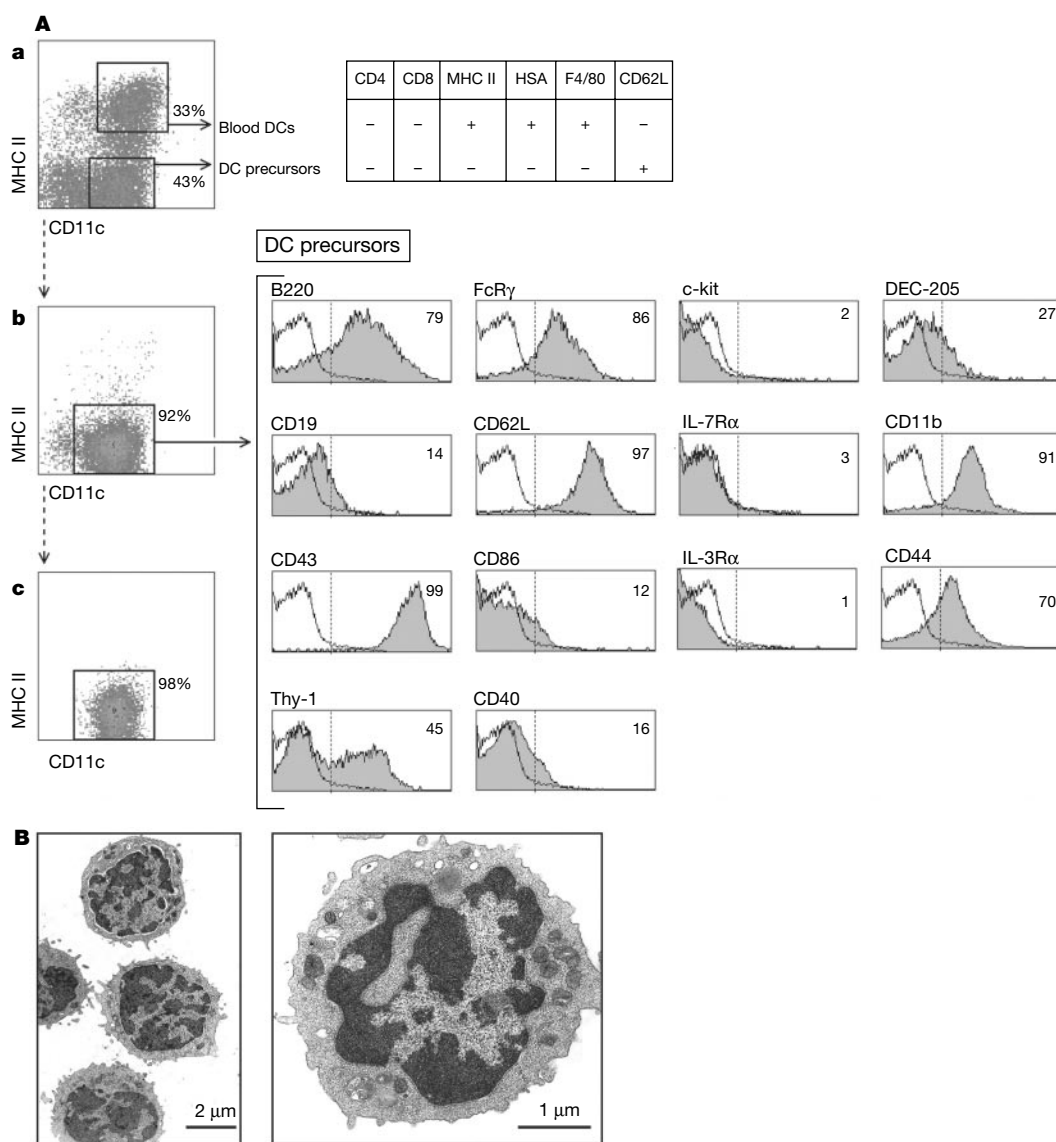
Analysis of the migration to the spleen of the CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors, performed after transfer of CFSE-labelled precursors revealed that around 1% of the injected CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors (that is, 5 × 10<sup>3</sup> cells) were detected in the spleen along the first week (not shown; CFSE, carboxyfluorescein diacetate succinimidyl ester). Therefore, as the total Ly5.2<sup>+</sup> DCs generated were approximately 30 × 10<sup>3</sup> by day 7 (Fig. 2b), these data indicate that the number of CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors had increased by a factor of six during the first week. In order to give added support to these data on the expansion of CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors, their *in vivo* proliferation rate was assessed after CFSE-labelling of purified Ly5.2<sup>+</sup> CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors and transfer into irradiated mice. As illustrated in Fig. 2c, the CFSE profile of CD11c<sup>+</sup> cells at day 7 after precursor transfer, after gating for Ly5.2<sup>+</sup> cells, showed that CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors underwent up to 3 cell divisions after the first week, in accordance with the sixfold expansion in numbers observed by day 7. These data reflect the strong expansion of CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors occurring during DC reconstitution, and are in agreement with those reported for DC generation from lymphoid and myeloid progenitors<sup>4</sup>.

We have recently found a CD11c<sup>+</sup> MHC-II<sup>+</sup> B220<sup>+</sup> CD40<sup>−</sup> DC subpopulation, which we will refer to here as B220<sup>+</sup>-DCs, characterized by its low stimulatory potential and capacity to produce interferon (IFN)-α after viral stimulation<sup>8</sup>. Therefore we analysed non-B220-depleted splenic DC-enriched fractions for the presence of donor-type B220<sup>+</sup>-DCs. Figure 2d shows the analysis of DC reconstitution 10 days after transfer of CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors, performed in parallel on B220<sup>+</sup> or B220<sup>−</sup> splenic DC-enriched fractions. Both donor-type CD11c<sup>+</sup> B220<sup>+</sup> DCs and CD11c<sup>+</sup> B220<sup>−</sup> DCs (the latter corresponding to CD8<sup>−</sup> plus CD8<sup>+</sup> DCs) were produced. The kinetics of B220<sup>+</sup>-DC production from DC-precursors paralleled that of B220<sup>−</sup> DCs; that is, donor-type B220<sup>+</sup>-DCs were detected at day 7 and peaked by day 14 (Fig. 2b). The donor-type B220<sup>+</sup>:B220<sup>−</sup> DC ratio and cell number at day 7, 10 and 14 was similar to that obtained from bone marrow precursors (not shown), demonstrating that differentiation from CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors included the full reconstitution of B220<sup>+</sup>-DCs. In order to determine the capacity of B220<sup>+</sup>-DCs generated from CD11c<sup>+</sup> MHC-II<sup>−</sup> precursors to produce IFN-α, as described for B220<sup>+</sup>-DCs from control mice<sup>8</sup>, FACS-sorted Ly5.2<sup>+</sup> B220<sup>+</sup>-DCs were analysed for IFN-α production (FACS, fluorescence-activated cell

sorting). Incubation with Sendai virus induced IFN- $\alpha$  mRNA expression by donor-type B220<sup>+</sup>-DCs, as assessed by RT-PCR (PCR after reverse transcription of RNA) analysis (Fig. 2e).

Analysis of the thymus, spleen, bone marrow and blood revealed that no T cells, B cells, monocytes, macrophages or granulocytes were generated from CD11c<sup>+</sup> MHC-II<sup>+</sup> precursors (Fig. 3A). Furthermore, neither B cell, nor myeloid cell, nor NK cell formation was observed after culture of CD11c<sup>+</sup> MHC-II<sup>+</sup> precursors under specific conditions for the differentiation of these cell lineages (not shown; see Methods for details). The percentage of DC reconstitution from CD11c<sup>+</sup> MHC-II<sup>+</sup> precursors was significantly lower for thymic DCs than for splenic DCs (0.1% versus 7% at day 7; not shown), suggesting a reduced thymus homing capacity of CD11c<sup>+</sup> MHC-II<sup>+</sup> precursors. In summary, blood CD11c<sup>+</sup> MHC-II<sup>+</sup> cells represent a DC-committed precursor population, hereafter named DC-precursors, with capacity to fully reconstitute all splenic DC subpopulations, including CD8<sup>+</sup> DCs, CD8<sup>+</sup> DCs and B220<sup>+</sup>-DCs, but devoid of lymphoid/myeloid differentiation potential.

Although it has been shown that DCs can be reconstituted from either lymphoid or myeloid precursors<sup>2,3</sup>, the DC differentiation pathway under physiological conditions, and consequently the functional relevance of DC-precursors in this process, remain to be clarified. In an attempt to trace the lineage derivation of DC-precursors, we analysed their expression of the transcription factors Pax-5 and SCL which are expressed by lymphoid and myeloid precursors in a mutually exclusive way<sup>9,10</sup>. RT-PCR analysis of FACS-sorted DC-precursors revealed that they expressed neither Pax-5 nor SCL-mRNA (Fig. 3B). In addition, flow cytometry analysis of the expression of the cytokine receptor IL-7R $\alpha$ , defining lymphoid-lineage commitment<sup>11</sup>, revealed that DC-precursors were IL-7R $\alpha$ <sup>+</sup> (Fig. 1A, b). These data might indicate that DC-precursors are located in a downstream position in the lymphoid and/or myeloid differentiation pathway involving Pax-5, IL-7R $\alpha$  and SCL downregulation; this would be in agreement with the non-expression by DC-precursors of the c-kit receptor (Fig. 1A, b) defining early haematopoietic precursors<sup>12</sup>. Alternatively, a multipotential



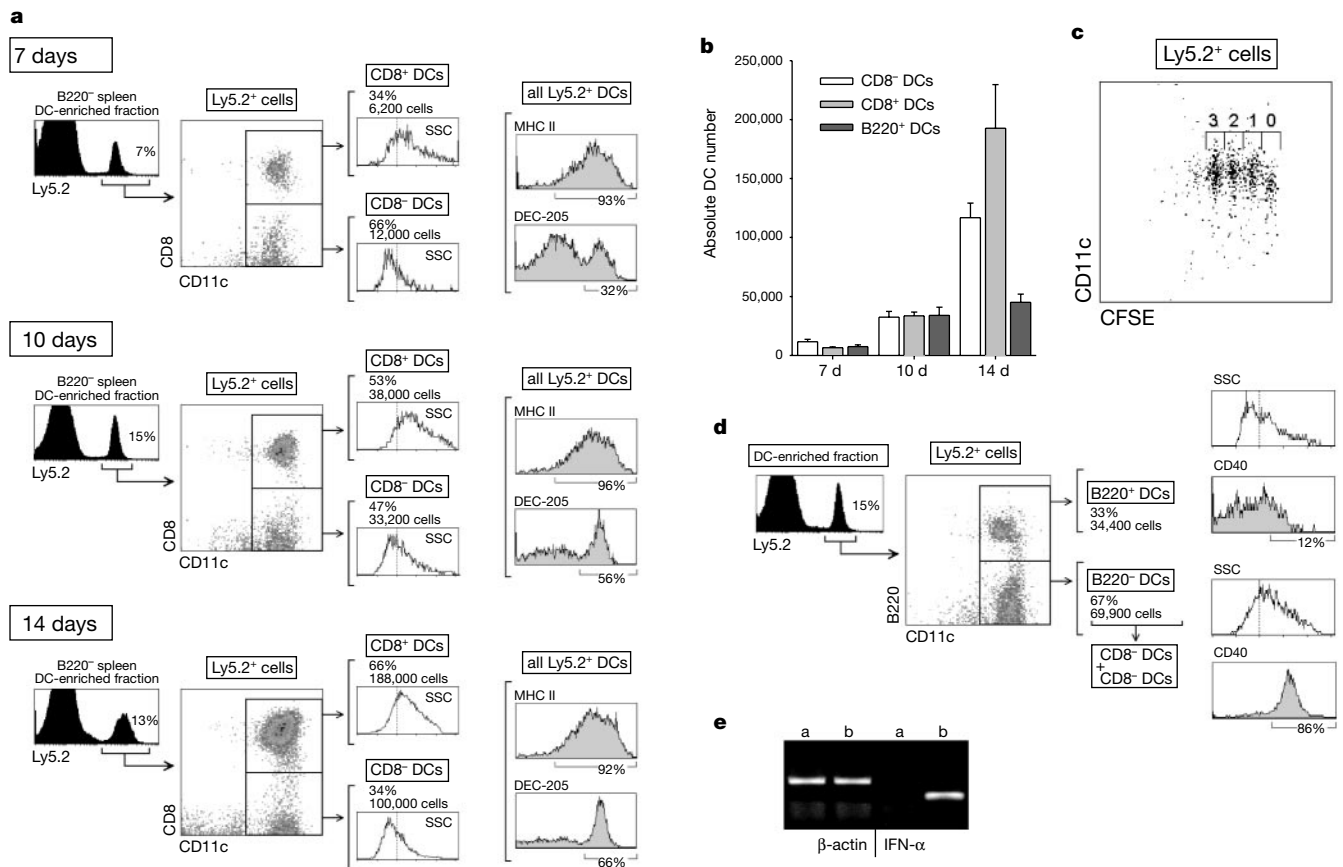
**Figure 1** Identification of DC-precursors in mouse peripheral blood. **A**, **a**, Definition of CD11c<sup>+</sup> MHC-II<sup>+</sup> and CD11c<sup>+</sup> MHC-II<sup>+</sup> blood cells performed on blood DC-enriched fractions. **b**, Phenotype of DC-precursors after isolation by a magnetic bead depletion method yielding a purity >90%. The percentage of cells with a fluorescence intensity over

the dotted vertical lines, corresponding to the background staining shown in the B220 histogram (white profile), is indicated. **c**, Purification of DC precursors by magnetic cell sorting (MACS), after magnetic bead depletion (purity >95%). **B**, Electron micrographs of FACS-sorted DC precursors (DC-precursor size:  $4.5 \pm 0.2 \mu$ m,  $n = 4$ ).

lymphoid/myeloid progenitor<sup>13,14</sup> may generate DC-precursors independently from lymphoid and myeloid progenitors. Whether physiologically DCs are generated as the result of a dual contribution of lymphoid and myeloid progenitors as proposed<sup>3,4</sup>, and/or independently from DC-precursors, has to be unravelled. (A model of DC development relying on a common DC differentiation pathway from DC-precursors is presented as Supplementary Information).

To explore whether DC-precursors were endowed with type-I IFN producing capacity upon viral stimulation, characterizing human plasmacytoid cells<sup>15,16</sup>, purified DC-precursors were analysed for the production of IFN $\alpha$  after incubation with Sendai virus. As assessed by RT-PCR analysis, no IFN $\alpha$ -mRNA was detected in DC-precursors after viral stimulation (Fig. 3B). Consequently, DC-precursors are not functionally equivalent to plasmacytoid cells, whose mouse counterpart has not been characterized yet, though their role could be fulfilled at least partly by B220<sup>+</sup>-DCs, endowed with IFN- $\alpha$  production capacity, and displaying specific phenotypic characteristics<sup>8</sup>. (Comparative phenotype profiles of DC-precursors versus B220<sup>+</sup> DCs were as follows. DC-precursors are MHC-II<sup>+</sup> CD4<sup>+</sup> CD8<sup>+</sup> HSA<sup>+</sup> CD62L<sup>+</sup> CD11b<sup>+</sup> whereas B220<sup>+</sup> DCs are MHC-II<sup>+</sup> CD4<sup>+</sup> CD8<sup>+</sup> HSA<sup>+</sup> CD62L<sup>+</sup> CD11b<sup>+</sup> (ref. 8).

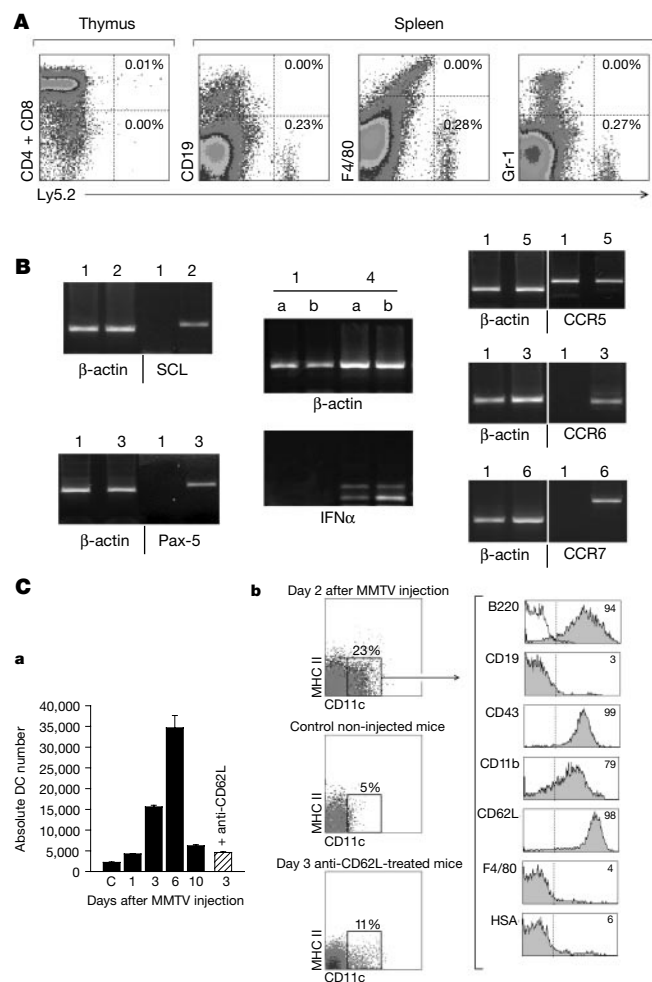
Finally, we investigated whether DC-precursors were involved in viral infection-induced inflammatory responses involving intensive DC recruitment. We analysed the DC-precursor population in the popliteal lymph nodes after injection of mouse mammary tumour virus (MMTV). MMTV induced a massive increase in lymph node DCs (Fig. 3C, a) that peaked by day 6 and declined afterwards, as described for the overall response to MMTV<sup>17</sup>. Two days after MMTV injection, CD11c<sup>int</sup> MHC-II<sup>+</sup> cells with an identical phenotype to blood DC-precursors were identified (Fig. 3C, b). These cells were found from day 2 to day 5 after injection, being undetectable after day 6 (not shown). These results suggest that MMTV-induced increase in DCs occurred by recruitment of blood DC-precursors and subsequent intranodal DC differentiation. In support of this concept, no significant increase in lymph node DCs (by day 3; Fig. 3C, a) or in CD11c<sup>int</sup> MHC-II<sup>+</sup> cells (by day 2; Fig. 3C, b) was induced by MMTV after anti-CD62L treatment, indicating that anti-CD62L blocked DC-precursor homing. It is important to note that anti-CD62L could not have blocked mature DC homing to the lymph node because blood DCs were CD62L-negative (Fig. 1A, a). Regarding DC-precursor expansion during DC differentiation accompanying MMTV infection, around  $7 \times 10^3$  ( $7,300 \pm 1,600$ ,  $n = 3$ ) DC-precursor-like CD11c<sup>int</sup> MHC-II<sup>+</sup> cells were detected at



**Figure 2** DC reconstitution by blood DC-precursors. **a**, CD8<sup>-</sup> and CD8<sup>+</sup> DC reconstitution from DC-precursors. Contour plots correspond to the CD11c versus CD8 expression of Ly5.2<sup>+</sup> cells analysed on B220<sup>-</sup> DC-enriched splenic fractions. The percentage and absolute number of CD8<sup>-</sup> and CD8<sup>+</sup> DCs is indicated over the side scatter histograms. Grey histograms show MHC-II and DEC-205 expression by donor-type splenic DCs. These results are representative of four independent experiments with similar results. **b**, Splenic CD8<sup>-</sup>, CD8<sup>+</sup> and B220<sup>+</sup>-DC reconstitution potential of DC-precursors. Histograms show the absolute number of reconstituted Ly5.2<sup>+</sup> CD8<sup>-</sup>, CD8<sup>+</sup> or B220<sup>+</sup>-DCs at day 7, 10 and 14 after DC-precursor transfer. Data represent the mean  $\pm$  s.d. ( $n = 4$ ). **c**, *In vivo* expansion of DC-precursors during DC reconstitution. Contour plots show CFSE-labelling

versus CD11c expression after gating for Ly5.2<sup>+</sup> cells on splenic DC-enriched fractions, analysed at day 7 after DC-precursor transfer. The number of cell divisions undergone by DC-precursors is indicated. **d**, B220<sup>+</sup>-DC generation by DC-precursors analysed on DC-enriched splenic fractions avoiding depletion with anti-B220, 10 days after DC-precursor transfer. The percentage and absolute number, as well as the side scatter and CD40 expression (grey profiles) of donor-type B220<sup>+</sup>, and B220<sup>-</sup>-DCs (which include prototypic splenic CD8<sup>-</sup> or CD8<sup>+</sup> DCs) are indicated. These results are representative of 3 independent experiments with similar results. **e**, IFN- $\alpha$  production by donor-type B220<sup>+</sup>-DCs. RT-PCR analysis of IFN- $\alpha$  mRNA expression by FACS-sorted Ly5.2<sup>+</sup> B220<sup>+</sup>-DCs after culture in the absence (a) or the presence (b) of Sendai virus.





**Figure 3** Links with myeloid and lymphoid lineages and homing potential of DC-precursors. **A**, Non-DC reconstitution potential of DC-precursors. Analysis performed on total thymus or spleen cell suspensions, 14 days after DC-precursor transfer into irradiated recipients, showing that no T cells, B cells, monocytes, macrophages, or granulocytes were generated from DC-precursors. (Note that the minute population of Ly5.2<sup>+</sup> CD8<sup>+</sup> thymic cells corresponds to CD8<sup>+</sup> DCs, as confirmed by our data on thymic DC reconstitution from DC-precursors, and that no monocytes or macrophages were found among DC-precursor progeny, since no donor-type Gr-1<sup>+</sup> and/or F4/80<sup>high</sup> cells were found, Ly5.2<sup>+</sup> F4/80<sup>low</sup> cells corresponding to reconstituted CD11c<sup>+</sup> splenic DCs. **B**, RT-PCR analysis of SCL, Pax-5, IFN- $\alpha$ , CCR5, CCR6 and CCR7 mRNA expression by DC precursors. 1: DC-precursors; 2: bone marrow cells; 3: splenic B cells; 4: peritoneal macrophages; 5: CD8<sup>+</sup> DCs; 6: CD40-stimulated Langerhans cells; a, b: 18-hour-culture in the absence (a) or presence (b) of Sendai virus. **C**, DC-precursor recruitment during MMTV infection. **a**, Absolute lymph node DC number at the indicated time points after MMTV injection. Data represent the mean  $\pm$  s.d. ( $n = 3$ ). **b**, Characterization of CD11c<sup>+</sup> MHC-II<sup>+</sup> DC-precursors in the lymph node at day 2 after virus injection. These results are representative of three experiments with similar results.

day 2 after MMTV injection, and around  $3.5 \times 10^4$  DCs were found at day 6, suggesting a 5-fold expansion. In order to further explore the mechanisms involved in the migration of DC-precursors, these were analysed for the expression of the chemokine receptors CCR5, CCR6 and CCR7 known to participate in DC homing and migration<sup>1</sup>. Neither CCR6- nor CCR7-mRNA was expressed by DC-precursors (Fig. 3B). However, DC-precursors expressed CCR5, which has been claimed to be involved in immature DC migration<sup>18</sup>, and could therefore control DC-precursor homing to lymphoid organs. □

## Methods

### Blood DC-precursor isolation

DC-precursors were isolated from lysis-buffer-treated heparinized blood by depletion (at a 7:1 bead-to-cell ratio) with magnetic beads (Dyna) coated with anti-rat immunoglobulins (Igs) and mouse Igs, after incubation with monoclonal antibodies anti-CD3, CD4, CD8 $\alpha$ , MHC-II, HSA, F4/80 and Gr-1 (purity >90%; Fig. 1A, b). Highly purified DC-precursor populations with a purity >98% were then obtained by magnetic cell sorting (MACS; Miltenyi Biotec) after incubation with anti-CD11c-conjugated microbeads (Fig. 1A, c), or fluorescence-activated sorting (FACS).

### Reconstitution experiments with DC-precursors

$5 \times 10^5$  DC-precursors purified from C57 BL/Ka Ly 5.2 donor mice were injected intravenously into  $\gamma$ -irradiated (7 Gy) C57 BL/6 Ly 5.1 Pep<sup>3b</sup> recipient mice, along with  $4 \times 10^4$  Ly 5.1 bone marrow cells to ensure survival of recipients. Donor-type DCs were analysed on B220<sup>+</sup> or B220<sup>+</sup> splenic DC-enriched fractions obtained from spleen low-density fractions, after magnetic bead depletion of cells expressing B220, CD3 or Gr-1 (B220<sup>+</sup> fraction) or expressing Igs, CD3 or Gr-1 (B220<sup>+</sup> fraction). Analysis of DC-precursor expansion was performed after labelling of purified DC-precursors with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester (CFSE; Molecular Probes) at 1  $\mu$ M for 10 min at 37 °C.

### Flow cytometry

Analysis of blood DC-precursors and lymph node CD11c<sup>int</sup> MHC-II<sup>+</sup> cells were performed after staining with fluorescein (FITC)-conjugated anti-CD11c, phycoerythrin (PE)-conjugated anti-MHC-II and biotin-conjugated anti-B220, CD19, CD43, CD11b, Thy-1, CD62L, Fc $\gamma$ II/III, CD44, c-kit, IL-7R $\alpha$ , IL-3R $\alpha$ , CD40, CD86, DEC-205, F4/80 or HSA, followed by streptavidin-TriColor (Caltag Laboratories). Analysis of CD8<sup>+</sup> and CD8<sup>+</sup> DC reconstitution was performed after staining with FITC-conjugated anti-Ly5.2, PE-conjugated anti-CD8 $\alpha$  and biotin-conjugated anti-CD11c, followed by streptavidin-TriColor. Analysis of B220<sup>+</sup>-DC reconstitution was performed after staining with FITC-conjugated anti-Ly5.2, PE-conjugated anti-B220 and biotin-conjugated anti-CD11c, followed by streptavidin-TriColor. Analysis of Ly5.2<sup>+</sup> DC phenotype was performed after staining with FITC-conjugated anti-Ly5.2, PE-conjugated anti-CD11c and biotin-conjugated anti-MHC II or DEC-205, followed by streptavidin-TriColor. Analysis of CD40 expression by Ly5.2<sup>+</sup> B220<sup>+</sup> or B220<sup>+</sup>-DCs was performed after staining with FITC-conjugated anti-Ly5.2, PE-conjugated anti-B220 and biotin-conjugated anti-CD40 followed by streptavidin-TriColor. Analysis of CD40 expression by Ly5.2<sup>+</sup> CD8<sup>+</sup> or CD8<sup>+</sup> DCs was performed after staining with FITC-conjugated anti-Ly5.2, PE-conjugated anti-CD8 $\alpha$  and biotin-conjugated anti-CD40 followed by streptavidin-TriColor. Analysis of CFSE profile of DC-precursors after transfer was performed after staining with PE-conjugated anti-CD11c and biotin-conjugated anti-Ly5.2 followed by streptavidin-TriColor.

### Electron microscopy

FACS-sorted DC-precursors were fixed with 1% glutaraldehyde and 1% paraformaldehyde in 0.1 M pH 7.6 Sørensen phosphate buffer for 1 h at 4 °C, postfixed with 1% OsO<sub>4</sub> in the same buffer for 1 h at 4 °C, dehydrated in graded acetone solutions, and embedded in Embed-812 (Electron Microscopy Sciences).

### Detection of IFN $\alpha$ -mRNA

Purified DC-precursors, or FACS-sorted donor-type splenic B220<sup>+</sup>-DCs isolated at day 10 after transfer of DC-precursors, were cultured for 18 h alone or in the presence of  $10^6$  p.f.u. ml<sup>-1</sup> Sendai virus (strain Z) and analysed for the expression of IFN $\alpha$ -mRNA by RT-PCR.

### In vitro culture of DC-precursors

Generation of B cells and myeloid cells was assessed after culture of purified DC-precursors on S17 stromal cells at  $10^5$  cells ml<sup>-1</sup> in 96-well plates, in the presence of 1 ng ml<sup>-1</sup> murine IL-7, as described elsewhere<sup>19</sup>. Cultures were analysed at day 12 for the presence of myeloid cells and/or B cells on the basis of the expression of CD11c/Gr-1 and B220/CD19, respectively. NK cell generation was assessed after culture in the presence of  $10^3$  U ml<sup>-1</sup> murine IL-2, as described elsewhere<sup>20</sup>. Cultures were analysed at day 5 for the presence of cells expressing the pan-NK cell marker DX5.

### Analysis of DC-precursors during MMTV infection

BALB/c mice were given a 10  $\mu$ l injection ( $10^9$  virus particles) of MMTV (strain SW), in the hind footpad and the draining popliteal lymph nodes were analysed. Characterization of CD11c<sup>int</sup> MHC-II<sup>+</sup> cells was performed after magnetic bead depletion of cells positive for Igs, CD3 or Gr-1, by gating on cells with low forward and side scatter. Blocking of DC-precursor migration to the lymph nodes via high endothelial venules was achieved by intravenous injection of purified anti-CD62L antibodies (clone Mel-14) into mice which were injected 24 h before with MMTV(SW), and subsequently analysed by day 3.

### RT-PCR

mRNA was purified from FACS-sorted cells with magnetic beads (mRNA direct micro kit; Dynal), reverse transcribed and subjected to PCR amplification using the following primers: SCL (sense): 5'-GTT TTT GTC TAG AGT TTT TGA GCC-3'; SCL (anti-sense):

5'-GCA TGC TCA AGG CTG CTG ACT TGG-3'. Pax-5 (sense): 5'-CTACAGGCTCC GTGACGCAG-3'; Pax-5 (anti-sense): 5'-TCT CGG CCT GTG ACA ATA GG-3'. CCR5 (sense): 5'-ACT TGG GTG GTG GCT GTG TTT-3'; CCR5 (anti-sense): 5'-TTG TCT TGC TGG AAA ATT GAA-3'. CCR6 (sense): 5'-CTG CAG TTC GAA GTC ATC-3'; CCR6 (anti-sense): 5'-GTC ATC ACC ACC ATA ATG TTT-3'. CCR7 (sense): 5'-AGC ACC ATG GAC CCA GGG AAA CC-3'; CCR7 (anti-sense): 5'-CAG CAT CCA GAT GCC CAC A-3'. IFN- $\alpha$  (sense): 5'-TGT CTG ATG CAG CAG GTG G-3'; IFN- $\alpha$  (anti-sense): 5'-AAG ACA GGG CTC TCC AGA C-3'. PCR conditions were: SCL: 30 s/95 °C; 30 s/60 °C; 60 s/72 °C. Pax-5: 15 s/94 °C; 90 s/65 °C. CCR5: 15 s/94 °C; 30 s/57 °C; 60 s/72 °C. CCR6: 15 s/94 °C; 15 s/59 °C; 45 s/72 °C. CCR7: 15 s/94 °C; 90 s/68 °C. IFN- $\alpha$ : 40 s/94 °C; 40 s/62 °C; 60 s/72 °C. PCR products were: SCL: 425 bp; Pax-5: 439 bp; CCR5: 539 bp; CCR6: 320 bp; CCR7: 561 bp; IFN- $\alpha$ : 166 bp.

## Langerhans cell stimulation

CD40-stimulated Langerhans cells, used as positive control for CCR7 expression, were isolated from the ear epidermis and cultured for 48 h with 100  $\mu$ g ml<sup>-1</sup> anti-CD40 monoclonal antibody, as described elsewhere<sup>21</sup>.

Received 24 October; accepted 12 December 2001.

1. Banchereau, B. *et al.* Immunobiology of dendritic cells. *Annu. Rev. Immunol.* **18**, 767–811 (2000).
2. Martin, P. *et al.* Concept of lymphoid versus myeloid dendritic cell lineages revisited: both CD8 $\alpha^+$  and CD8 $\alpha^-$  dendritic cells are generated from CD4<sup>low</sup> lymphoid-committed precursors. *Blood* **96**, 2511–2519 (2000).
3. Traver, D. *et al.* Development of CD8 $\alpha$ -positive dendritic cells from a common myeloid progenitor. *Science* **290**, 2152–2154 (2000).
4. Manz, M. G., Traver, D., Miyamoto, T., Weissman, I. L. & Akashi, K. Dendritic cell potentials of early lymphoid and myeloid progenitors. *Blood* **97**, 3333–3341 (2001).
5. Anjuère, F. *et al.* Definition of dendritic cell subpopulations present in the spleen, Peyer's patches, lymph nodes and skin of the mouse. *Blood* **93**, 590–598 (1999).
6. Martínez del Hoyo, G., Martin, P., Fernández Arias, C., Rodríguez-Marín, A. & Ardavin, C. CD8 $\alpha^+$  dendritic cells originate from the CD8 $\alpha^-$  dendritic cell subset by a maturation process involving CD8 $\alpha$ , DEC-205 and CD24 upregulation. *Blood* **99**, 999–1004 (2002).
7. Ardavin, C., Wu, L., Li, C. L. & Shortman, K. Thymic dendritic cells and T cells develop simultaneously in the thymus from a common precursor population. *Nature* **362**, 761–763 (1993).
8. Martin, P. *et al.* Characterization of a new subpopulation of mouse CD8 $\alpha^+$  B220<sup>+</sup> dendritic cells with tolerogenic potential. *Blood* (submitted).
9. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
10. Mebius, R. E. *et al.* The fetal liver counterpart of adult common lymphoid progenitors gives rise to all lymphoid lineages, CD45<sup>+</sup>CD4<sup>+</sup>CD3<sup>+</sup> cells, as well as macrophages. *J. Immunol.* **166**, 6593–6601 (2001).
11. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
12. Lyman, S. D. & Jacobsen, S. E. W. *c-kit* ligand and flt3 ligand: Stem/progenitor cell factors with overlapping yet distinct activities. *Blood* **91**, 1101–1134 (1998).
13. Singh, H. Gene targeting reveals a hierarchy of transcription factors regulating specification of lymphoid cell fates. *Curr. Opin. Immunol.* **8**, 160–165 (1996).
14. Mojica, M. P. *et al.* Phenotypic distinction and functional characterization of pro-B cells in adult mouse bone marrow. *J. Immunol.* **166**, 3042–3051 (2001).
15. Siegal, F. P. *et al.* The nature of the principal type I interferon-producing cells in human blood. *Science* **284**, 1835–1837 (1999).
16. Cella, M. *et al.* Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon. *Nature Med.* **5**, 919–923 (1999).
17. Ardavin, C. *et al.* B cell response after MMTV infection: extrafollicular plasmablasts represent the main infected population and can transmit viral infection. *J. Immunol.* **162**, 2538–2545 (1999).
18. Sozzani, S. *et al.* Cutting edge: Differential regulation of chemokine receptors during dendritic cell maturation: A model for their trafficking properties. *J. Immunol.* **161**, 1083–1086 (1998).
19. Tudor, K.-S. R., Payne, K. J., Yamashita, Y. & Kincade, P. W. Functional assessment of precursors from murine bone marrow suggests a sequence of early B lineage differentiation events. *Immunity* **12**, 335–345 (2000).
20. Rolink, A. *et al.* A subpopulation of B220<sup>+</sup> cells in murine bone marrow does not express CD19 and contains natural killer cell progenitors. *J. Exp. Med.* **183**, 187–194 (1996).
21. Anjuère, F., Martínez del Hoyo, G., Martin, P. & Ardavin, C. Langerhans cells acquire a CD8 $\alpha^+$  dendritic cell phenotype on maturation by CD40 ligation. *J. Leukocyte Biol.* **67**, 206–209 (2000).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

This work was supported by the European Commission, Comunidad de Madrid and Ministerio de Ciencia y Tecnología of Spain. We thank A. Rolink for the anti-CD40 hybridoma FGK45, K. Akashi for advice on SCL detection, G. Márquez for CCR6 and CCR7 primers, D.F. Tough for IFN $\alpha$  primer sequences, D. Kolakofsky for Sendai virus, H. Acha-Orbea for MMTV, and A. Rodríguez-Marín and V. Parrillas for discussions.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A. (e-mail: [ardavin@bio.ucm.es](mailto:ardavin@bio.ucm.es)).

# A blue-light-activated adenylyl cyclase mediates photoavoidance in *Euglena gracilis*

Mineo Iseki<sup>†‡</sup>, Shigeru Matsunaga<sup>§</sup>, Akio Murakami<sup>||</sup>, Kaoru Ohno<sup>\*</sup>, Kiyoshi Shiga<sup>†</sup>, Kazuichi Yoshida<sup>\*</sup>, Michizo Sugai<sup>#</sup>, Tetsuo Takahashi<sup>✉</sup>, Terumitsu Hori<sup>§</sup> & Masakatsu Watanabe<sup>\*††</sup>

<sup>\*</sup> National Institute for Basic Biology, Okazaki, Aichi, 444-8585 Japan

<sup>†</sup> Bio-oriented Technology Research Advancement Institution, 3-18-19, Toranomon, Minato-ku, Tokyo, 105-0001 Japan

<sup>‡</sup> Center for Gene Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8602 Japan

<sup>§</sup> Institute of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8572 Japan

<sup>||</sup> Kobe University Research Center for Inland Seas, Iwaya, Awaji, Hyogo, 656-2401 Japan

<sup>¶</sup> Department of Physiology, Kumamoto University School of Medicine, Honjo, Kumamoto, Kumamoto, 860-8556 Japan

<sup>#</sup> Department of Biology, Faculty of Science, Toyama University, Toyama, Toyama, 930-8555 Japan

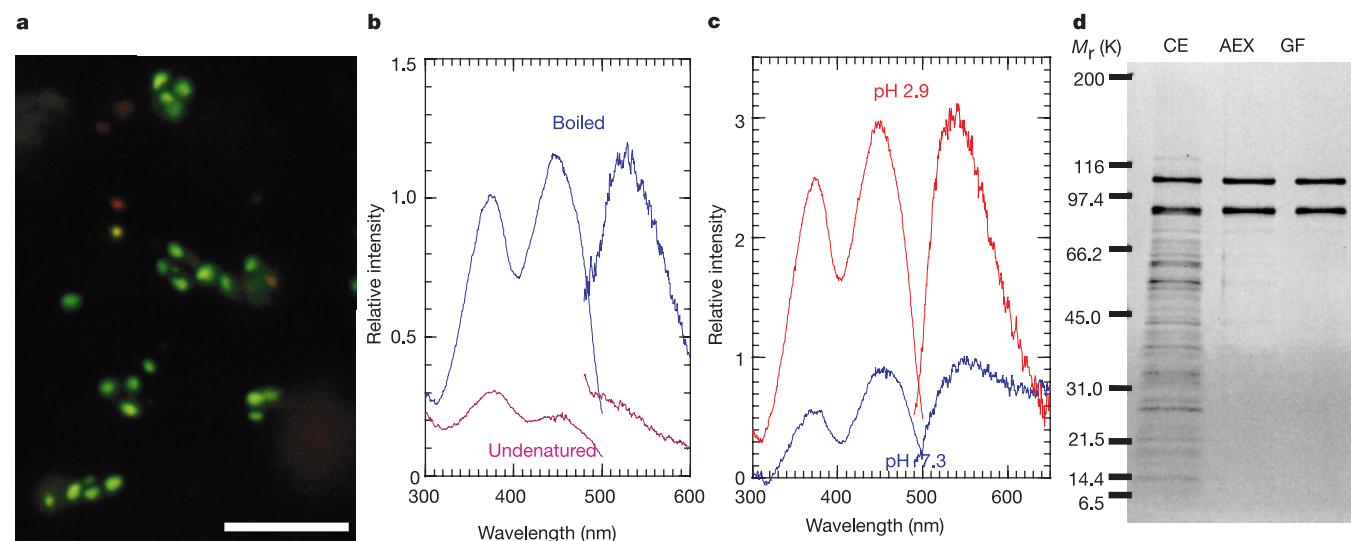
<sup>✉</sup> School of Pharmaceutical Sciences, Toho University, 2-2-1, Miyama, Funabashi, Chiba, 274-8510 Japan

<sup>††</sup> Department of Photoscience, School of Advanced Sciences, Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa, 240-0193 Japan

Blue light regulates processes such as the development of plants and fungi and the behaviour of microbes<sup>1,2</sup>. Two types of blue-light receptor flavoprotein have been identified: cryptochromes, which have partial similarity to photolyases<sup>3,4</sup>, and phototropins, which are photoregulated protein kinases<sup>5,6</sup>. The former have also been found in animals with evidence of essential roles in circadian rhythms<sup>7,8</sup>. *Euglena gracilis*, a unicellular flagellate, abruptly changes its swimming direction after a sudden increase or decrease in incident blue light intensity, that is, step-up or step-down photophobic responses, resulting in photoavoidance or photoaccumulation, respectively<sup>9</sup>. Although these photobehaviours of *Euglena* have been studied for a century<sup>10</sup>, the photoreceptor molecules mediating them have remained unknown<sup>9</sup>. Here we report the discovery and biochemical characterization of a new type of blue-light receptor flavoprotein, photoactivated adenylyl cyclase, in the photoreceptor organelle of *Euglena gracilis*, with molecular genetic evidence that it mediates the step-up photophobic response.

Action spectroscopy has indicated that one or more flavins are chromophores of the photoreceptor molecules for photophobic responses in *Euglena*<sup>11,12</sup>. The paraflagellar body (PFB), near the base of its flagellum and in the vicinity of the stigma, is considered as a photosensing organelle for photomovements<sup>9</sup> (see Fig. 4A). Its bright green autofluorescence supports the hypothesis that flavo-proteins localized in the PFB might act as the photoreceptor molecule<sup>13,14</sup>. To examine this hypothesis we isolated PFBs and purified the flavoproteins from them.

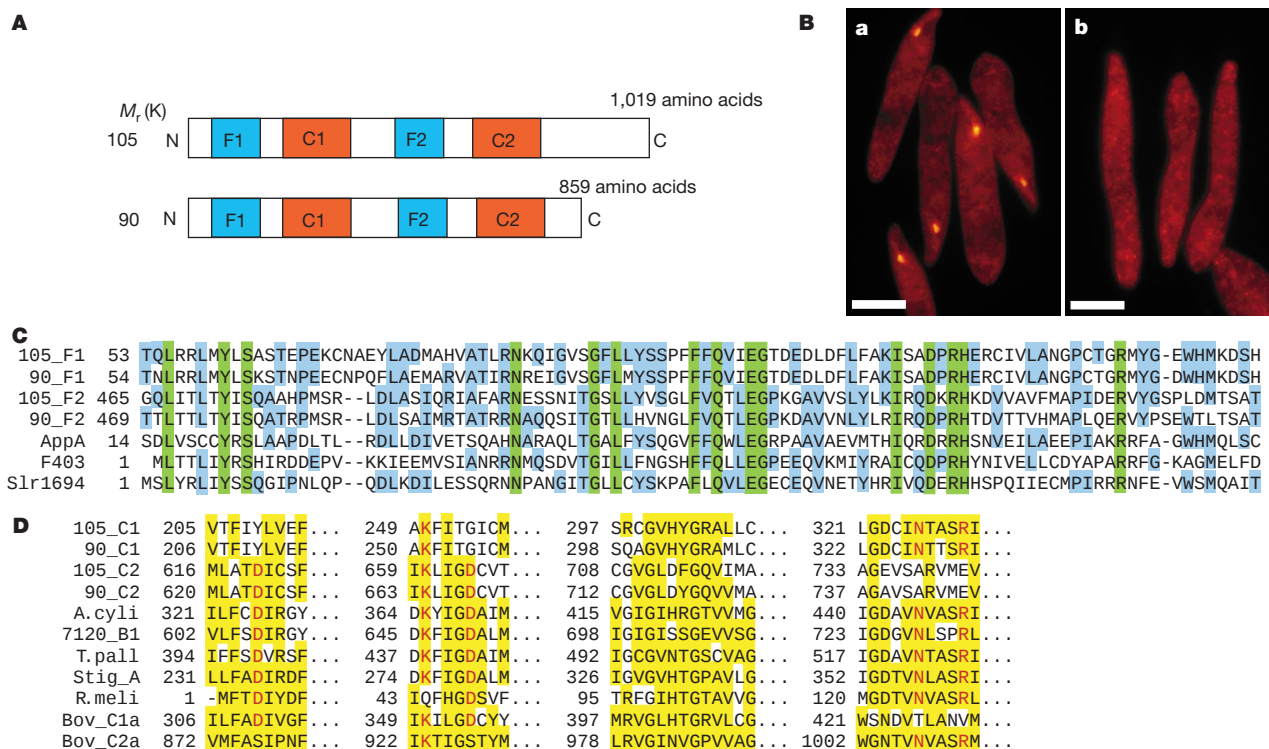
The isolated PFBs (Fig. 1a) were solubilized by sonication in detergent and the flavoprotein was purified by anion-exchange chromatography, followed by gel filtration. The fraction of apparent molecular mass ~400,000 ( $M_r$  ~400K) showed the highest intensity of fluorescence at 530 nm after boiling to release noncovalently bound chromophores (see Fig. 3b), with excitation (peaking at 370 and 450 nm) and emission (peaking at 530 nm) spectra characteristic of flavins (Fig. 1b). The fluorescence excitation spectrum of the undenatured  $M_r$  400K fraction also showed two peaks at 370 and 450 nm (Fig. 1b), consistent with the action spectra for photophobic responses of *Euglena*<sup>12</sup>. The fact that the fluorescence of the undenatured  $M_r$  400K fraction was much weaker than



**Figure 1** Purification of the flavoprotein from isolated paraflagellar bodies.

**a**, Fluorescence micrograph of isolated PFBs (green fluorescent particles) observed under excitation by blue–violet light. Scale bar, 10  $\mu$ m. **b**, Fluorescence excitation (left, 530 nm emission) and emission (right, 370 nm excitation) spectra of the boiled and undenatured  $M_r$  400K fractions (25  $\mu$ g ml<sup>-1</sup> protein). **c**, Fluorescence excitation (left, 530 nm emission) and emission (right, 370 nm excitation) spectra of the boiled  $M_r$  400K fraction at different

pH values (pH 2.9, adjusted with citrate buffer; pH 7.3, the same pH as the gel-filtration buffer at room temperature.) **d**, SDS–PAGE of PFB proteins at consecutive purification steps: CE, crude extract of PFB; AEX, the 530-nm fluorescence peak fraction after anion-exchange chromatography; GF, the  $M_r$  400K fraction from gel filtration. A 5–20% gradient gel, stained with silver.



**Figure 2** Sequencing of the  $M_r$  105K and 90K polypeptides. **A**, Structural features of the two polypeptides. Four characteristic regions are shown as blue (F1, F2) and red (C1, C2) boxes. The DDBJ/EMBL/GenBank accession numbers for the  $M_r$  105K and 90K polypeptides are AB031225 and AB031226, respectively. **B**, Indirect immunofluorescence staining of *Euglena* cells with polyclonal antibodies against the C terminus of the  $M_r$  105K polypeptide (**a**) or with preimmune antisera as a negative control (**b**). Scale bar, 10  $\mu$ m. **C**, Sequence alignment of F1 and F2 with AppA flavin (FAD)-binding domain and its homologues. Regions of similarity (blue) and identity (green) are highlighted. Alignment and similarity identification were performed with GCG software. AppA, *Rhodobacter*

*sphaeroides* redox regulator (L42555); F403, *E. coli* hypothetical protein (AE000215); Slr1694, *Synechocystis* sp. PCC6803 hypothetical protein (D90913). **D**, Partial sequence alignment of C1 and C2 with catalytic domains from class III adenylyl cyclases. Amino acids consistent with the four consensus sequences of class III adenylyl cyclases<sup>19</sup> (yellow) and those identical with important amino acids for catalysis in mammalian adenylyl cyclases<sup>30</sup> (red) are highlighted. A. cyl., *Anabaena cylindrica* (D55650); 7120\_B1, *Anabaena* sp. PCC7120 CyaB1 (D89623); T. pall., *Treponema pallidum* (AE001224); Stig\_A, *Stigmatella aurantiaca* CyaA (AJ223796); R. meli., *Rhizobium meliloti* (M35096); Bov\_C1a and Bov\_C2a, bovine Type I (M25579).



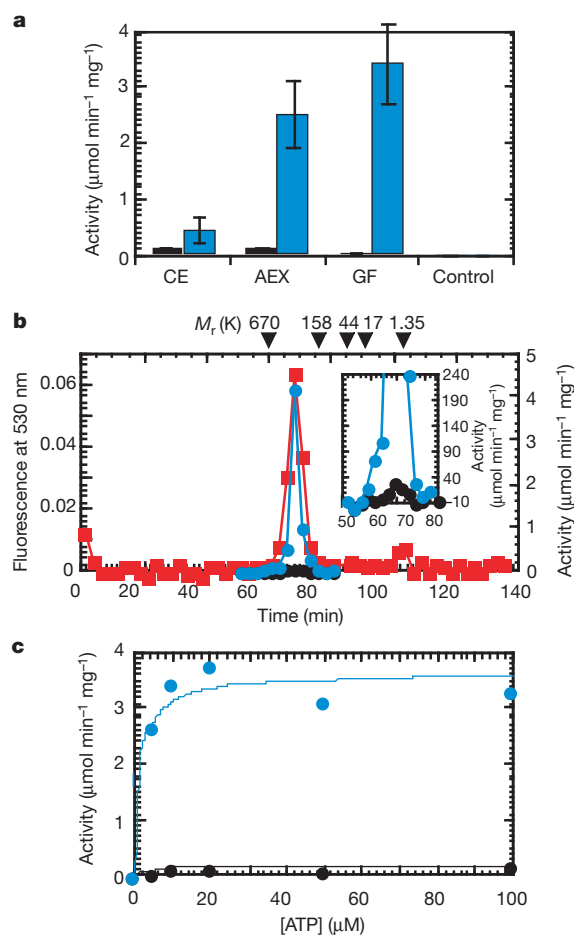
that of the detached flavin chromophore indicates that the fluorescence of protein-bound flavin was quenched by the surrounding protein matrix. The fluorescence intensity of the boiled fraction at pH 7.3 was ~30% of that at pH 2.9 (Fig. 1c). FAD shows a similar decrease in fluorescence intensity at higher pH owing to the formation of an intramolecular complex between the flavin ring and the adenine ring, whereas riboflavin and flavin mononucleotide (FMN) do not<sup>15</sup>. In addition, thin-layer chromatography with butan-1-ol/acetic acid/water (4:1:5) as solvent revealed that the mobility of the chromophore was close to that of the authentic FAD:  $R_F$  values for the chromophore, FAD, FMN and riboflavin were 0.09, 0.11, 0.18 and 0.45, respectively. We therefore postulate that the chromophore is most probably FAD, although strict structural information has yet to be determined.

Separation of the  $M_r$  400K fraction by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) showed two equally stained protein bands corresponding to  $M_r$  values of 105K and 90K (Fig. 1d). This indicates that the  $M_r$  400K flavoprotein consists of two distinct subunits of similar sizes, indicating that it has a  $(105K)_2(90K)_2$  heterotetrameric structure. We determined amino-terminal amino-acid sequences of the polypeptides and synthesized degenerate primers for the amplification of complementary DNAs by polymerase chain reaction (PCR). Nucleotide sequences of the cDNAs were determined by direct sequencing of the PCR products and extended by 5' and 3' rapid amplification of cDNA ends (RACE) methods. The deduced amino-acid sequences of the  $M_r$  105K and 90K polypeptides revealed that they are composed of 1,019 and 859 amino acids with calculated  $M_r$  values of 111,976 and 94,215, respectively, and are homologous over the region where they overlap: the sequence identity over the 800 amino acids of the N-terminal portion is 75% (Fig. 2A). Hydropathy profiles of the full-length polypeptides indicated no apparent transmembrane helices. Immunoblot analysis with polyclonal antibodies against oligopeptides that had been synthesized in accordance with the carboxy-terminal sequences of the polypeptides each showed specific staining of  $M_r$  105K and 90K bands, indicating that the sequenced polypeptides were indeed expressed in *Euglena* cells (data not shown). Localization of the  $M_r$  105K polypeptide in the PFB was confirmed by immunostaining (Fig. 2B) with the polyclonal antibody. In both sequences we found four characteristic regions: a pair of homologous regions (F1 and F2) are separated by one (C1) of the other homologous pairs C1 and C2 (Fig. 2A). F1 and F2 show homology with the N-terminal regions of a redox regulator of photosystem formation in *Rhodobacter* (AppA)<sup>16</sup>, hypothetical proteins found in genome sequences of *Escherichia coli* (F403)<sup>17</sup> and *Synechocystis* (Slr1694)<sup>18</sup>. Because AppA and F403 bind FAD noncovalently at an apparent molar ratio of 1:1, the homologous regions in these proteins are considered to be flavin (FAD)-binding domains<sup>16</sup>. The sequence identity between F1, F2 and the flavin-binding domains of AppA, F403 and Slr1694 are almost the same (27–38%) for any combination of these, and most of the conserved amino acids within AppA, F403 and Slr1694 are also conserved in F1 and F2 (Fig. 2C). The apparent molar ratio of FAD to the  $M_r$  400K protein of *Euglena*, estimated from flavin fluorescence and by assay of the protein by the Bradford method, varied from 2.5:1 (FAD to protein) to 7:1 in several experiments. Taking into account the sequence similarity and the limited accuracy of this estimation, we tentatively assume that the  $M_r$  105K and 90K polypeptides each bind two FAD molecules per polypeptide, at F1 and F2.

The other characteristic regions, C1 and C2, quite unexpectedly, show homology with catalytic domains from class III adenylyl cyclases<sup>19</sup>, especially bacterial ones. Although the sequence identity over the full length of the adenylyl cyclase catalytic domains is not very high (27% between C1 and *Treponema* adenylyl cyclase<sup>20</sup>, and 28% between C2 and *Anabaena* adenylyl cyclase<sup>21</sup>), four consensus amino-acid sequences in class III adenylyl cyclases<sup>19</sup> are well conserved in C1 and C2 (Fig. 2D). To examine whether the  $M_r$  400K

protein is actually an adenylyl cyclase, we determined the adenylyl cyclase activity at each purification step. Whereas the adenylyl cyclase activity in each sample was slight but significant in darkness, it was elevated up to 80-fold under blue light (Fig. 3a). In the gel filtration, the adenylyl cyclase activity showed a peak that coincided exactly with the peak of 530-nm fluorescence intensity (Fig. 3b), indicating that the activity arose from the flavoprotein. The blue-light-activated adenylyl cyclase activity of the  $M_r$  400K fraction was saturated at 50  $\mu$ M ATP (Fig. 3c), with a  $V_{max}$  (3,500 pmol min<sup>-1</sup> mg<sup>-1</sup>) and an estimated  $K_m$  (0.5  $\mu$ M) for ATP comparable to those of cyanobacterial adenylyl cyclase<sup>22</sup>. From these results, we conclude that the flavoprotein isolated from PFBs is an adenylyl cyclase with its activity regulated by blue light. We therefore designate the flavoprotein as photoactivated adenylyl cyclase (PAC) and its subunits, the  $M_r$  105K and 90K polypeptides, as PAC $\alpha$  and PAC $\beta$ , respectively.

To examine whether PAC actually mediates the photophobic responses of *Euglena*, we suppressed the expression of PAC $\alpha$  and PAC $\beta$  by RNA-mediated interference (RNAi)<sup>23,24</sup>: double-stranded RNAs (dsRNAs) corresponding to N-terminal portions of the



**Figure 3** Adenylyl cyclase activity of the flavoprotein purified from PFBs. **a**, Adenylyl cyclase activity of proteins at consecutive purification steps: CE, crude extract of PFB; AEX, the 530-nm fluorescence peak fraction after anion-exchange chromatography; GF, the  $M_r$  400K fraction from gel filtration; control, gel-filtration buffer as a control. Means and s.e.m. from three measurements in darkness (black) or under blue light (blue) at 10  $\mu$ mol m<sup>-2</sup> s<sup>-1</sup> are shown. **b**, Gel-filtration profile showing 530-nm fluorescence and adenylyl cyclase activity. Fluorescence emission at 530 nm excited by 370 nm light (red) and adenylyl cyclase activity measured in darkness (black) or under blue light (blue) are shown. The inset shows an enlarged view of adenylyl cyclase activities around the  $M_r$  400K peak. **c**, ATP concentration dependence of adenylyl cyclase activity of the  $M_r$  400K fraction measured in darkness (black) or under blue light (blue).

subunits were introduced into *Euglena* cells by electroporation. The dsRNAs-containing cells grew as well as control cells that had been electroporated with buffer only. The PFB in most of the dsRNAs-containing cells disappeared as observed with Nomarski optics and epifluorescence microscopy (Fig. 4A). This phenotype was maintained even after several cycles of cell division, indicating that the effect of dsRNAs was epigenetically inheritable. Northern blots of the dsRNAs-containing cells showed no detectable signal when probed with probes specific for PAC $\alpha$  or PAC $\beta$  (Fig. 4B), indicating that endogenous messenger RNAs of PAC $\alpha$  and PAC $\beta$  are degraded after introduction of the dsRNAs. These results mean that the RNAi system worked well in suppressing the expression of PAC $\alpha$  and PAC $\beta$  in *Euglena* cells.

Observation by a computerized video motion analyser<sup>12</sup> showed that the step-up photophobic response disappeared in the dsRNAs-containing cells even at a light intensity sufficient to cause saturation of the response in the control cells, whereas the step-down response of the dsRNAs-containing cells was essentially the same as that of the control cells (Fig. 4C). This means that the introduction of the dsRNAs impaired only the photosensing system for the step-

up photophobic response without affecting the machinery needed to exhibit phobic responses. From these results we conclude that PAC is the major constituent of the PFB and it acts as the photoreceptor for the step-up photophobic response in *Euglena*, whereas it is not involved in the step-down photophobic response.

An intracellular increase of cyclic AMP after the light activation of PAC in the PFB would affect flagellar motility as in metazoan sperm<sup>25</sup> and *Chlamydomonas*<sup>26</sup>, thus bringing about the step-up photophobic response of *Euglena*. PAC is a unique protein that can act both as a photoreceptor and as an effector to catalyse cAMP synthesis, in contrast with G-protein-coupled receptor systems, in which three different proteins act sequentially to modulate the level of cyclic nucleotide<sup>27</sup>. This simple mechanism indicates that PAC might be a promising tool in cell biology and biotechnology for the photomanipulation of the intracellular cAMP level in heterologous cell systems. □

## Methods

### Isolation of PFBs

*Euglena* cells grown for 10 d ( $\sim 5 \times 10^5$  cells ml<sup>-1</sup>) under conditions described previously<sup>12</sup> were harvested from a 2.7-litre culture and washed in a buffer (50 mM potassium phosphate buffer pH 7.0, 0.5 mM EDTA). The cells were resuspended in the same buffer with 0.3 M sucrose added, and disrupted in a nitrogen cavitation vessel (8.5 MPa, 2 min, Parr Cell Disruption Bomb 4635). After centrifugation (400g, 15 min, 4 °C) to remove remaining cells and large cell debris, smaller particles containing PFBs were pelleted (6,500g, 15 min, 4 °C) and subjected to discontinuous density-gradient centrifugation (0.8, 1.5, 1.75 and 2.0 M sucrose in the washing buffer, 113,000g, 1 h, 4 °C). The boundary between 1.75 and 2.0 M sucrose was collected and mixed with an equal volume of the washing buffer and filtered (Nuclepore, 3  $\mu$ m; Costar). The filtrate was centrifuged (6,500g, 10 min, 4 °C) to obtain a final pellet containing  $(1-3) \times 10^7$  PFBs.

### Purification of the flavoprotein

The PFB-containing pellet was resuspended in a lysis buffer (40 mM Tris-HCl pH 8.0 at 4 °C, 0.2% *n*-dodecyl- $\beta$ -D-maltoside, 10 mM dithiothreitol, 1 mM phenylmethylsulphonyl fluoride (PMSF)) and sonicated. After centrifugation of the lysate (12,000g, 10 min, 4 °C), the supernatant was concentrated and applied to an anion-exchange column (POROS HQ/M; PerSeptive Biosystems) equilibrated with a starting buffer (40 mM Tris-HCl pH 8.0 at 4 °C, 0.02% *n*-dodecyl- $\beta$ -D-maltoside, 1 mM dithiothreitol, 0.1 mM PMSF) and eluted with a linear gradient of NaCl (0–0.5 M in starting buffer). The eluate was collected in 78 fractions; part of each fraction was boiled for 5 min before fluorescence measurement. Fluorescence emission spectra of the boiled fractions were measured with a spectrofluorometer (Hitachi model 850) to identify the fraction containing the largest amount of flavoprotein. The identified fraction and the adjacent two fractions were combined, concentrated and then applied to a gel-filtration column (Superose 6HR10/30; Amersham Pharmacia Biotech) equilibrated with a gel-filtration buffer (150 mM NaCl in starting buffer). The eluate was collected in 55 fractions; part of each fraction was used for fluorometric analysis as described above and the remainder of each fraction was used for SDS-PAGE and adenylyl cyclase activity assay. All manipulations were performed under dim red light ( $<0.2 \mu\text{mol m}^{-2} \text{s}^{-1}$ , 590–720 nm).

### Sequencing of PAC $\alpha$ and PAC $\beta$

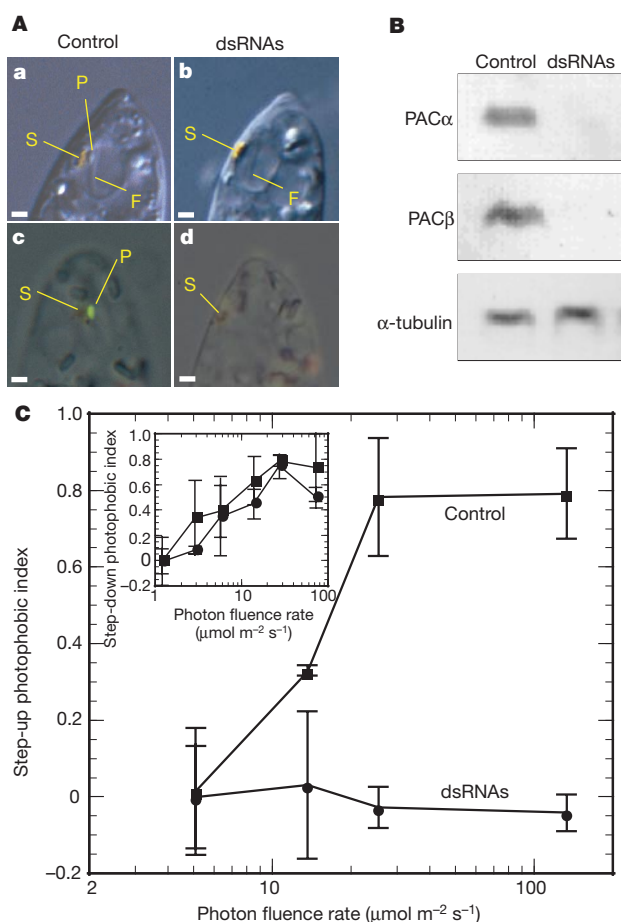
Partial amino-acid sequences of the N termini of PAC $\alpha$  and PAC $\beta$  were determined with a protein sequencer (PE Applied Biosystems model 494). On the basis of the sequences obtained (MYILVWKEGQQTIRTFQDLLEEXGQFQTASNIT for PAC $\alpha$  and XYILVWKKGGQKIKFTHTLDEAAQFKAASNIDE for PAC $\beta$ ; one-letter amino acid codes), degenerate primers were designed for use in the PCR amplification of cDNAs synthesized from poly(A)<sup>+</sup> RNAs purified from 7-day-old *Euglena* cells. Fragments ( $\sim 100$  bp) were obtained from the PCR and sequenced with an automated DNA sequencer (PRISM 310; PE Applied Biosystems). Unknown regions downstream of the 100-bp fragments were amplified by 3' RACE and the resultant PCR products (3.6 kilobases for PAC $\alpha$  and 3.0 kilobases for PAC $\beta$ ) were sequenced directly. The upstream regions of the 100-bp fragments were amplified by 5' RACE. Translation start points of the cDNAs were confirmed by finding the characteristic leader sequence of *Euglena* mRNAs<sup>28</sup> in the 5' RACE products. For the PCR primers used, see Supplementary Information.

### Antibodies

Antibodies were purified by a protein A affinity column from antisera of rabbits immunized with keyhole-limpet-haemocyanin-conjugated oligopeptides (CASTDLKGSYKYEYH for PAC $\alpha$  and CHGDSGRRNSAQGKRS for PAC $\beta$ ). The secondary antibody for immunostaining was purchased from Sigma (T5268).

### Adenylyl cyclase assay

The adenylyl cyclase activities of PAC were measured over 10 min in an assay buffer (50 mM Tris-HCl, pH 7.5, 1 mM MnCl<sub>2</sub>, 1 mM dithiothreitol, 0.1% bovine serum



**Figure 4** Suppression of gene expression of PAC by RNAi. **A**, Photomicrographs of anterior portion of *Euglena* cells electroporated with buffer only (**a, c**) or dsRNAs of PAC $\alpha$  and PAC $\beta$  (**b, d**). **a, b**, Nomarski optics; **c, d**, fluorescence images, under excitation by blue-violet light, superimposed with bright-field images of the same cell. S, stigma; F, flagellum; P, PFB. Scale bar, 1  $\mu$ m. **B**, Northern blots of total RNA extracted from cells to which dsRNAs had been introduced and from control cells probed with probes specific for PAC $\alpha$ , PAC $\beta$  and  $\alpha$ -tubulin. **C**, Fluence-rate-response curves for the step-up photophobic response of the dsRNAs-containing cells (circles) and control cells (squares) at 450 nm, measured as described previously<sup>12</sup>. Fluence-rate-response curves for the step-down photophobic response are also shown (inset). Means and s.d. from three measurements are shown.

albumin) containing 100  $\mu$ M ATP (except in Fig. 3c) at 27 °C. During the 10 min, the reaction mixture was either kept in darkness (under dim red light as mentioned above) or irradiated continuously with blue light (450 nm, 10  $\mu$ mol m<sup>-2</sup> s<sup>-1</sup>) provided by the Okazaki Large Spectrograph<sup>29</sup>. cAMP was assayed with a BIOTRAK EIA system (Amersham Pharmacia Biotech).

## RNAi experiments

Synthesis and electroporation of the dsRNAs were performed in accordance with the methods for *Trypanosoma*<sup>24</sup>, with several modifications (for details see Supplementary Information). Because a dark-grown culture of *Euglena* clearly shows the step-up photophobic response, the electroporated cells were maintained in darkness for measurement of this response. In contrast, light-grown cells, which exhibit a clear step-down response, were used for measurement of that response. Both dark-grown and light-grown cells showed a disappearance of the PFB after introduction of the dsRNAs. Measurements of the step-up and the step-down responses were made as described previously<sup>12</sup>. Stimulus light (450 nm) was provided by the Okazaki Large Spectrograph<sup>29</sup>.

Received 28 September 2001; accepted 7 January 2002.

1. Senger, H. & Schmidt, W. in *Photomorphogenesis in Plants* 2nd edn (eds Kendrick, R. E. & Kronenberg, G. H. M.) 301–325 (Kluwer Academic, Dordrecht, 1994).
2. Watanabe, M. in *Handbook of Organic Photochemistry and Photobiology* (eds Horspool, W. & Song, P.-S.) 1276–1288 (CRC Press, Boca Raton, 1995).
3. Ahmad, M. & Cashmore, A. R. *HY4* gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. *Nature* **366**, 162–166 (1993).
4. Cashmore, A. R., Jarillo, J. A., Wu, Y.-J. & Liu, D. Cryptochromes: blue light receptors for plants and animals. *Science* **284**, 760–765 (1999).
5. Huala, E. *et al.* *Arabidopsis* NPH1: a protein kinase with a putative redox-sensing domain. *Science* **278**, 2120–2123 (1997).
6. Briggs, W. R. *et al.* The phototropin family of photoreceptors. *Plant Cell* **13**, 993–997 (2001).
7. Emery, P. *et al.* CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95**, 669–679 (1998).
8. van der Horst, G. T. J. *et al.* Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398**, 627–630 (1999).
9. Lebert, M. in *Comprehensive Series in Photosciences 1: Photomovement* (eds Häder, D.-P. & Jori, G.) 297–341 (Elsevier, Amsterdam, 2001).
10. Buder, J. Zur Kenntnis der phototaktischen Richtungsbewegungen. *Jb. Wiss. Bot.* **58**, 105–220 (1919).
11. Barghigiani, C., Colombetti, G., Franchini, B. & Lenci, F. Photobehavior of *Euglena gracilis*: action spectrum for the step-down photophobic response of individual cells. *Photochem. Photobiol.* **29**, 1015–1019 (1979).
12. Matsunaga, S. *et al.* Discovery of signaling effect of UV-B/C light in the extended UV-A/blue-type action spectra for step-down and step-up photophobic responses in the unicellular flagellate alga *Euglena gracilis*. *Protoplasma* **201**, 45–52 (1998).
13. Benedetti, P. A. & Lenci, F. *In vivo* microspectrofluorometry of photoreceptor pigments in *Euglena gracilis*. *Photochem. Photobiol.* **26**, 315–318 (1977).
14. Brodhun, B. & Häder, D.-P. Photoreceptor proteins and pigments in the paraflagellar body of the flagellate *Euglena gracilis*. *Photochem. Photobiol.* **52**, 865–871 (1990).
15. Bessey, O. A., Lowry, O. H. & Love, R. H. The fluorometric measurement of the nucleotides of riboflavin and their concentration in tissues. *J. Biol. Chem.* **180**, 755–769 (1949).
16. Gomelsky, M. & Kaplan, S. AppA, a redox regulator of photosystem formation in *Rhodospirillum rubrum*, is a flavoprotein. Identification of a novel FAD binding domain. *J. Biol. Chem.* **273**, 35319–35325 (1998).
17. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
18. Kaneko, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
19. Danchin, A. Phylogeny of adenylate cyclases. *Adv. Sec. Mess. Phosphoprot. Res.* **27**, 109–162 (1993).
20. Fraser, C. M. *et al.* Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**, 375–388 (1998).
21. Katayama, M., Wada, Y. & Ohmori, M. Molecular cloning of the cyanobacterial adenylate cyclase gene from the filamentous cyanobacterium *Anabaena cylindrica*. *J. Bacteriol.* **177**, 3873–3878 (1995).
22. Kasahara, M., Yashiro, K., Sakamoto, T. & Ohmori, M. The *Spirulina platensis* adenylate cyclase gene, *cyaC*, encodes a novel signal transduction protein. *Plant Cell Physiol.* **38**, 828–836 (1997).
23. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
24. Ngô, H., Tschudi, C., Gull, K. & Ullu, E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA* **95**, 14687–14692 (1998).
25. Darsoz, A., Labarca, P., Nishigaki, T. & Espinosa, F. Ion channels in sperm physiology. *Physiol. Rev.* **79**, 481–510 (1999).
26. Hasegawa, E., Hayashi, H., Asakura, S. & Kamiya, R. Stimulation of *in vitro* motility of *Chlamydomonas* axonemes by inhibition of cAMP-dependent phosphorylation. *Cell Motil. Cytoskel.* **8**, 302–311 (1987).
27. Torre, V., Ashmore, J. F., Lamb, T. D. & Menini, A. Transduction and adaptation in sensory receptor cells. *J. Neurosci.* **15**, 7757–7768 (1995).
28. Tessier, L.-H. *et al.* Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in *Euglena*. *EMBO J.* **10**, 2621–2625 (1991).
29. Watanabe, M. *et al.* Design and performance of the Okazaki Large Spectrograph for photobiological research. *Photochem. Photobiol.* **36**, 491–498 (1982).
30. Hurley, J. H. Structure, mechanism, and regulation of mammalian adenylate cyclase. *J. Biol. Chem.* **274**, 7599–7602 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

## Acknowledgements

We thank J. W. Hastings for critical reading of the manuscript, M. Ohmori and M. Ishiura for discussions, I. Nakashima and Y. Yamakawa for advice on subcellular fractionation, N. Murata for encouragement, Y. Makino for operation of the protein sequencer at NIBB Center for Analytical Instruments, and M. Itoh for technical assistance. Part of this work was done under the NIBB cooperative-research programme for the use of the Okazaki Large Spectrograph. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan, and by Grants for Ground Research for Space Utilization from the Japan Space Forum. M.I. thanks NIBB and BRAIN for a postdoctoral fellowship during part of this study.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W. (e-mail: machakou@nibb.ac.jp).

# Structure and dynamics of KH domains from FBP bound to single-stranded DNA

Demetrios T. Braddock<sup>\*†</sup>, John M. Louis<sup>\*</sup>, James L. Baber<sup>\*</sup>, David Levens<sup>†</sup> & G. Marius Clore<sup>\*</sup>

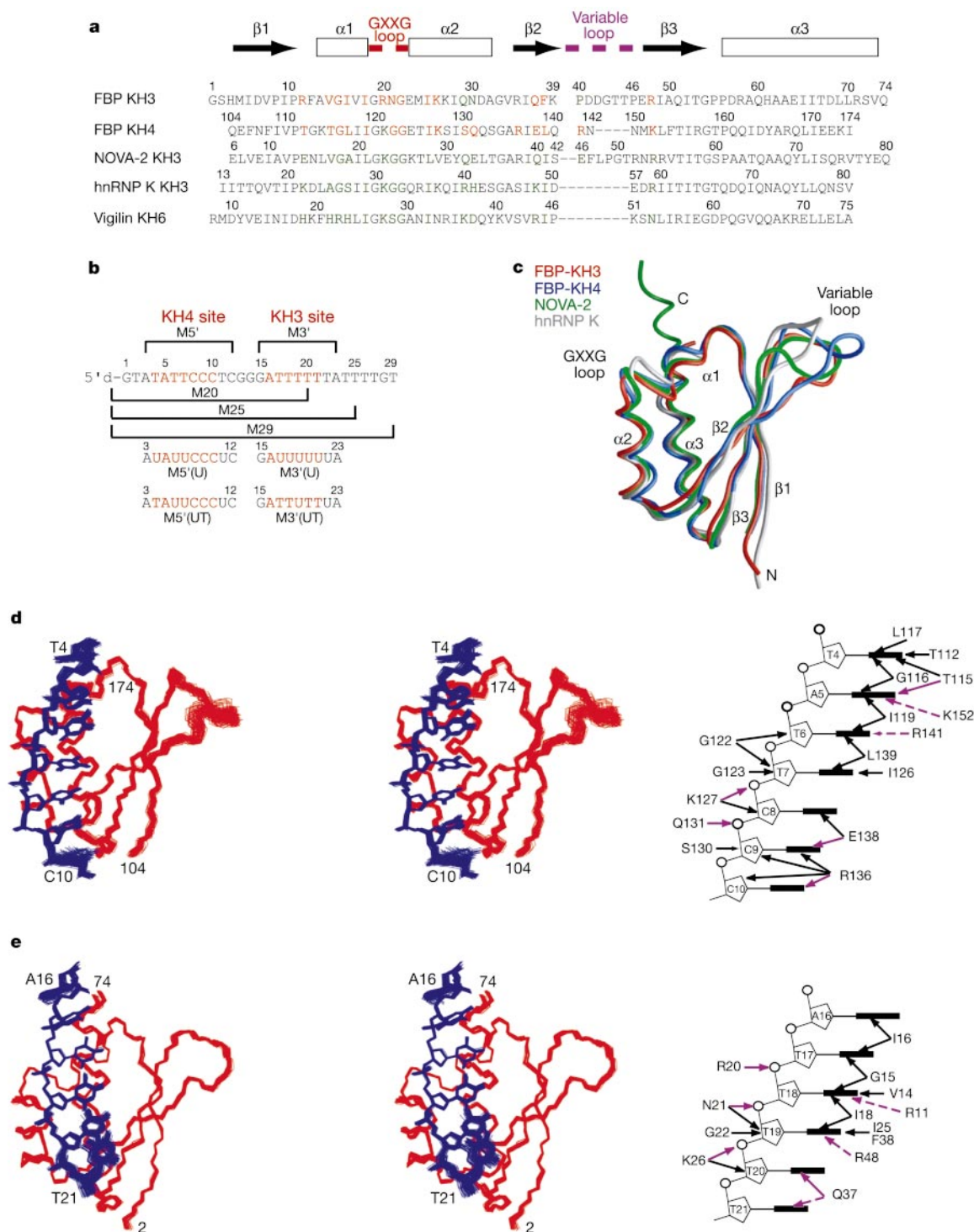
<sup>\*</sup> Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases; and <sup>†</sup> Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Gene regulation can be tightly controlled by recognition of DNA deformations that are induced by stress generated during transcription<sup>1–3</sup>. The KH domains of the FUSE-binding protein (FBP), a regulator of *c-myc* expression<sup>4,5</sup>, bind *in vivo* and *in vitro* to the single-stranded far-upstream element (FUSE), 1,500 base pairs upstream from the *c-myc* promoter<sup>4–6</sup>. FBP bound to FUSE acts through TFIID at the promoter<sup>4</sup>. Here we report the solution structure of a complex between the KH3 and KH4 domains of FBP and a 29-base single-stranded DNA from FUSE. The KH domains recognize two sites, 9–10 bases in length, separated by 5 bases, with KH4 bound to the 5' site and KH3 to the 3' site. The central portion of each site comprises a tetrad of sequence 5'-d-ATTC for KH4 and 5'-d-TTTT for KH3. Dynamics measurements show that the two KH domains bind as articulated modules to single-stranded DNA, providing a flexible framework with which to recognize transient, moving targets.

FBP contains four K homology (KH) repeats<sup>7</sup> separated by linkers of varying lengths<sup>5</sup>. The minimal single-stranded DNA (ssDNA) binding domain, as seen from gel electrophoresis<sup>6</sup>, comprises KH3 and KH4 (Fig. 1a). The target for KH3 and KH4 has been localized by sensitivity to the ssDNA selective agent permanganate to a 29-base stretch (M29 in Fig. 1b) in the non-coding strand of the FUSE element (–1525 to –1553 of the *c-myc* gene)<sup>6</sup>. We solved the three-dimensional structure for a complex of relative molecular mass ( $M_r$ ) of 30,000 comprising the KH3 and KH4 domains of human FBP (residues 278–447 of the full length sequence, referred to as FBP3/4 and numbered 5–174) bound to M29 ssDNA using multidimensional nuclear magnetic resonance<sup>8</sup>.

The complex of FBP3/4 with M29 ssDNA as well as with shorter ssDNAs, 25 (M25) and 20 (M20) bases long (Fig. 1b), is long-lived ( $K_{\text{diss}} \approx 10$  nM) and readily purified as a single complex by gel filtration. Intermolecular nuclear Overhauser experiments (NOEs) showed that KH4 and KH3 bound to bases 4–10 and 16–21, respectively, of M29 ssDNA. However, the spectrum of M29





**Figure 1** Structural analysis of the FBP3/4-ssDNA complex. **a**, Structure-based sequence alignment of the KH domains of FBP3/4, NOVA-2, hnRNP K and vigilin. Residues of the KH3 and KH4 domains of FBP3/4 that contact ssDNA are indicated in red; the equivalent residues in the other KH domains are coloured green. **b**, ssDNAs used in the current study. NMR analysis was carried out on FBP3/4 complexed to M29 and intermolecular NOE contacts were confirmed using complexes of the isolated KH4 and KH3 domains bound to M5' and M3' (UT), respectively. **c**, Backbone superposition of the KH domains of FBP3/4 (KH3 and KH4 in red and blue, respectively), NOVA-2 (ref. 9) (green) and hnRNP K<sup>10</sup> (grey). The C $\alpha$  root mean square (r.m.s.) differences range from 1.1 Å for FBP KH3 and KH4 versus NOVA-2 KH3<sup>9</sup> to 1.6 Å for FBP KH3 versus hnRNP K KH3 (ref. 10). **d**, **e**, Stereoviews showing best-fit superposition of the final 80 simulated annealing structures (protein backbone in red, DNA in blue) and a summary of the observed intermolecular contacts

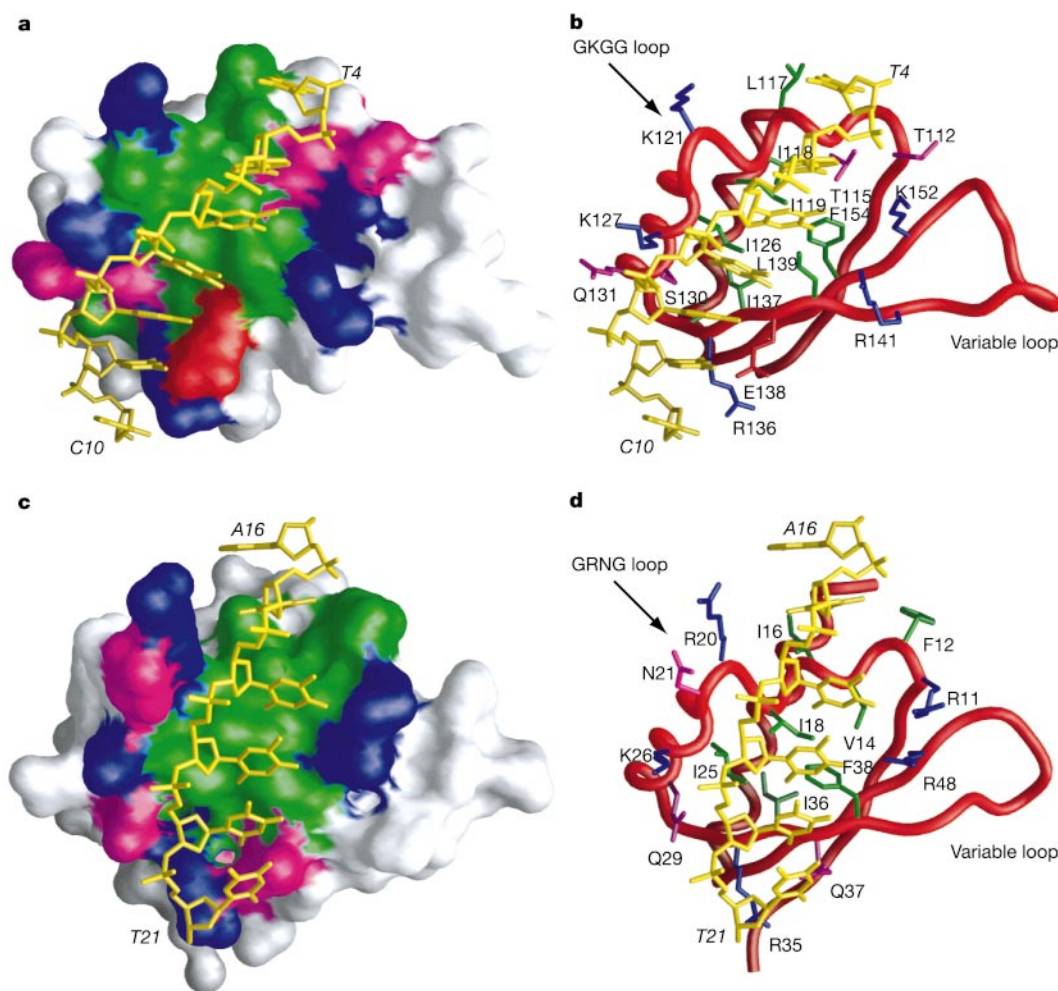
(with H-bonds and salt bridges represented by purple arrows; the dashed arrows indicate potential electrostatic interactions) for the KH4 and KH3 domains, respectively. The coordinate precision for the protein backbone plus DNA heavy atoms is 0.30 and 0.38 Å for the KH3 and KH4 halves of the complex, respectively. (The corresponding values for all heavy atoms are 0.64 and 0.70 Å, respectively.) The experimental NMR restraints for the KH3 and KH4 halves of the complex are as follows: 1,095/949 interproton distances (including 50/68 intermolecular contacts), 244/261 torsion angles, 33/36  $^3J_{\text{HN}}$  couplings, 120/121  $^{13}\text{C}\alpha/\beta$  shifts, and 61/61  $^1\text{D}_{\text{NH}}$ , 47/39  $^1\text{D}_{\text{NC}}$  and 46/40  $^2\text{D}_{\text{HNC}}$  dipolar couplings. There are no interproton distance or torsion angle violations  $>0.5$  Å and  $>5^\circ$ , respectively; the  $^1\text{D}_{\text{NH}}$  dipolar coupling  $R$ -factor<sup>30</sup> is  $<10\%$ . The percentage residues in the most favourable region of the Ramachandran map is 96% for KH3 and 90% for KH4. A complete table of structural statistics is provided in the Supplementary Information.

ssDNA displayed considerable overlap, particularly at the KH3-binding site. To resolve ambiguities and confirm the assignment of intermolecular NOEs, NMR data for the individual KH domains complexed with shorter ssDNAs (Fig. 1b) were also analysed: KH4 with M5' (bases 3–12 of M29) and KH3 with M3' (UT) (bases 15–23 of M29 with the T bases at positions 19 and 22 substituted by U to alleviate spectral degeneracy arising from the stretch of six T bases.  $K_{\text{diss}}$  for the binding of KH4 to M5' and KH3 to M3', M3'(U) and M3'(UT) is around 3  $\mu\text{M}$ ;  $K_{\text{diss}}$  for the binding of KH4 to M5'(U) and M5'(UT) is around 12  $\mu\text{M}$ . Thus, substitution of the methyl group of T by a hydrogen in U has no effect for the KH3-binding site and only a marginal effect in energetic terms for the KH4-binding site. The structure of the FBP3/4–M29 complex was determined from 3,153 experimental NMR restraints, including 118 intermolecular interproton distance restraints and 294 protein dipolar couplings. A best-fit superposition of the final 80 simulated annealing structures for the KH4–ssDNA and KH3–ssDNA domains is shown in Fig. 1d, e, respectively.

The KH fold<sup>9–11</sup> comprises three  $\alpha$ -helices packed onto a three-stranded antiparallel  $\beta$ -sheet in a  $\beta 1$ - $\alpha 1$ - $\alpha 2$ - $\beta 2$ - $\beta 3$ - $\alpha 3$  topology. The invariant GXXG sequence in the loop connecting helices 1 and 2 is characterized by positive  $\phi$  angles at positions 1, 3 and 4, with the

$\phi/\psi$  angles of residues 3 and 4 in a left-handed helical conformation. Strands  $\beta 2$  and  $\beta 3$  are linked by a loop of variable length (7–10 residues) and conformation. A best-fit superposition of the KH domains of FBP3/4, NOVA-2 (ref. 9) and hnRNP K<sup>10</sup> is shown in Fig. 1c, and the corresponding structure-based sequence alignments are provided in Fig. 1a.

Single-stranded DNA binds to a groove delineated by helices 1 and 2 and the GXXG loop on one side and strand  $\beta 2$  on the other (Fig. 2). The centre of the groove is hydrophobic and the edges are hydrophilic and charged (Fig. 2a, c). The sugar-phosphate backbone is directed towards the left-hand side of the binding site while the bases point towards the centre and right-hand side of the site in the view shown in Fig. 2. The narrow binding site ( $\sim 10$  Å) favours pyrimidines over purines, explaining why the KH3 and KH4 binding sites each contain a single purine, located at the 5' end. The B-like ssDNA is slightly underwound and extended, positioning the 5' end at the top of the binding site in close proximity to the GXXG motif (Fig. 2). All the sugar-phosphate backbone torsion angles of the ssDNA lie in a narrow range characteristic of right-handed DNA<sup>12</sup>, as evidenced by the restricted dispersion of the <sup>31</sup>P resonances (between 4 and 4.5 p.p.m.)<sup>13</sup>. The pattern of intra- and inter-nucleotide NOEs throughout is typical of right-handed DNA<sup>8</sup>.



**Figure 2** Structure of the FBP3/4–ssDNA complex. **a, b**, KH4 domain bound to bases 4–10 of M29. **c, d**, KH3 domain bound to bases 16–21 of M29. In **a** and **c** the protein is depicted as a molecular surface with hydrophobic, uncharged hydrophilic, positively charged and negatively charged residues comprising the ssDNA binding sites depicted in green, pink, blue and red, respectively; the ssDNA heavy atoms are shown in yellow. In

**b** and **d**, the protein backbone is depicted as a red tube, and the colouring for the protein side chains in the DNA-binding site and ssDNA are the same as in **a** and **c**. The coordinates of the KH3–ssDNA and KH4–ssDNA structures (restrained regularized mean) have been best-fitted to each other, and the same view is shown in all four panels. Nucleotide numbering is in italics.

Hydrophobic contacts to the central T bases of the KH3 (T18, T19) and KH4 (T6, T7) binding sites involve Ile 18, Ile 25 and Phe 38, and Ile 119, Ile 126 and Leu 139, respectively. The methyl groups of these thymines are directed towards solvent, explaining why substitution of T for U has little effect on binding affinity.

Recognition of the KH4 site involves a number of intermolecular H-bonds (Figs 1d and 2a, b): at the 5' end from the O $\gamma$  atom of Thr 115 to the N6 amino group of A5; at the 3' end from the carboxylate of Glu 138 to the N4 amide group of C9, and from the O2 and O4' atoms of the base and sugar of C10 to the guanidino group of Arg 136. Other electrostatic interactions involve the N $\epsilon$  amino group of Lys 152 and the N7 atom of A5, and the guanidino group of Arg 141 and the O4 atom of T6. Contacts to the sugar-phosphate backbone are afforded by Gly 122 and Gly 123 of the GXXG motif, and Lys 127, Ser 130 and Gln 131 of helix 2.

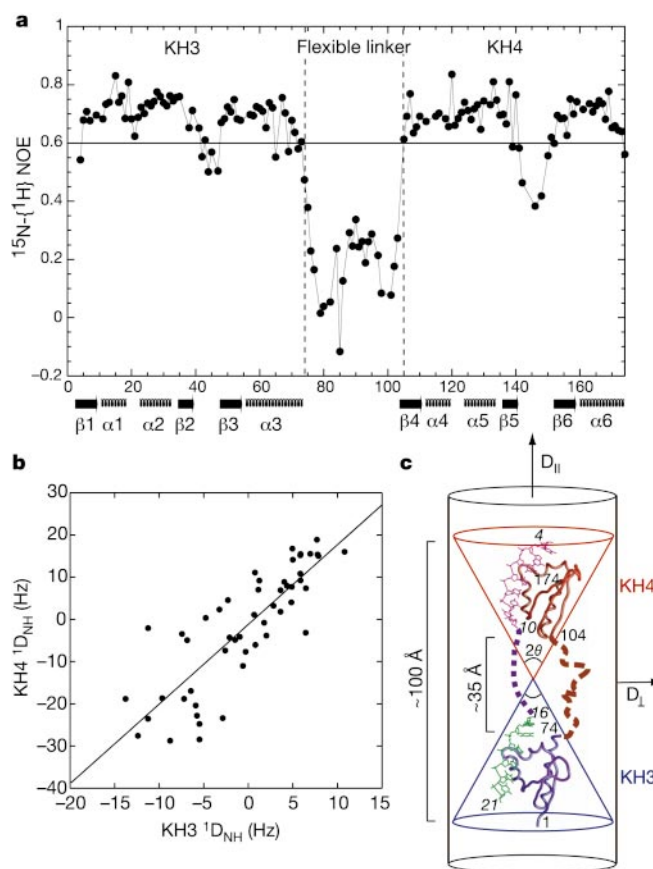
The KH3 domain recognizes a stretch of four T bases (T17–T20), the contacts with which are mainly hydrophobic (Figs 1e and 2c, d). At the 5' end there are H-bonds between the guanidino groups of Arg 11 and Arg 48 and the O4 atoms of T18 and T19, respectively; at the 3' end the side-chain amido group of Gln 37 is H-bonded to the O2 atom of T20, and possibly the O2 or N3 atoms of T21. Contacts to the sugar-phosphate backbone are afforded by Arg 20, Asn 21 and Gly 22 of the GXXG motif and Lys 26 of helix 2. Thr 115 in KH4, H-bonded to the base of A5, is replaced by Val 14 in KH3 which favours a T base at this position.

The KH domain presents a scaffold for nucleic acid binding that can be tuned for ssDNA<sup>5,6</sup> or RNA<sup>7,9</sup> specificity. The ssDNA-binding surfaces of the KH domains of FBP3/4 overlap extensively with that of the NOVA-2 KH3 domain that recognizes a stem-loop RNA<sup>9</sup>. There are, however, major differences. Extensive intermolecular contacts are made between ssDNA and residues in helix 2 and the amino-terminal end of strand  $\beta$ 2 in the FBP3/4 complex (Fig. 2), which are not seen in the NOVA-2 KH3–RNA complex where equivalent interactions are precluded owing to the presence of the wider double-stranded RNA stem. In contrast, in the NOVA-2 KH3–RNA complex contacts are made to both arms of the RNA hairpin and, as a consequence, the binding surface also includes the flexible loop (connecting strands  $\beta$ 2 and  $\beta$ 3) and the carboxy-terminal end of helix 3. The latter extends well beyond the protein core of NOVA-2 KH3 (Fig. 1c), presumably as a consequence of protein–RNA contacts and/or crystal packing. The core of the RNA-recognition sequence, 5'-UCAC<sup>9</sup>, is located and oriented similarly to the central portion of the KH4 (5'-d-ATTC) and KH3 (5'-d-TTTT) ssDNA recognition sequences, but the RNA and ssDNA tetrads have very different conformations. The bases of the ssDNA tetrads are stacked upon one another (Fig. 2). In contrast, the first three bases of the RNA tetrad are not stacked: the base at position 2 is oriented at right angles to the bases at positions 1 and 4, and concomitantly the conformation of the sugar-phosphate backbone departs markedly from that characteristic of right-handed RNA or DNA<sup>12</sup>. The unusual features of the RNA tetrad arise from a network of inter- and intra-molecular H-bonds involving the bases and 2'-OH groups of the ribose sugars.

The KH3 and KH4 domains of FBP (Fig. 1a) do not interact with one another in the FBP3/4–M29 ssDNA complex. They are connected by a glycine-rich, flexible 30-residue linker (Fig. 3a), and the DNA is bridged by 5 bases (bases 11–15) that do not display any contacts to protein. If the FBP3/4–M29 ssDNA complex were rigid, dipolar couplings measured for both halves of the complex in a dilute liquid crystalline medium, such as phage fd<sup>14</sup>, would be described by a single alignment tensor<sup>15,16</sup>. However, the magnitude of the alignment tensor<sup>15</sup> for the KH3 domain is half that of the KH4 domain in the complex (Fig. 3b), indicative of significant interdomain motion<sup>16</sup> (see Supplementary Information for further details), which was characterized by model free analysis of <sup>15</sup>N relaxation data using the extended Lipari–Szabo model<sup>17,18</sup>.

Assuming an axially symmetric diffusion tensor<sup>18</sup>, the data yield values of 21.5 and 1.85 ns for the effective overall rotational correlation time and diffusion anisotropy, respectively, of the whole complex. The timescale and magnitude of the interdomain motions are described by an effective slow internal correlation time ( $\tau_s$ ) of 4.1 ns for KH3 and 3.6 ns for KH4, and an order parameter  $S_s^2$  of 0.67 for KH3 and 0.70 for KH4, which correspond to the two KH domains wobbling independently in cones with semi-angles of about 30° (Fig. 3c). Alignment of the principal components of the diffusion tensors of the two domains indicates that the average orientation of the two domains is parallel (average interhelical angle between the third helix of each domain is about 1°, see Fig. 3c).

The interdomain mobility of the FBP3/4–M29 ssDNA complex is



**Figure 3** Interdomain motion in the FBP3/4–M29 ssDNA complex. **a**, <sup>15</sup>N-{<sup>1</sup>H} NOE values as a function of residue number. The low NOE values for the 30-residue linker indicate high flexibility for this segment of the polypeptide chain. **b**, Correlation of <sup>1</sup>D<sub>NH</sub> dipolar couplings of structurally equivalent residues measured for the KH4 and KH3 domains in the FBP3/4–M29 complex. The correlation coefficient is 0.83, as expected for very similar structures, but the magnitude of the alignment tensor<sup>15</sup> for the N–H vectors in the KH3 domain (–7.2 Hz) is half that in the KH4 domain (–14.5 Hz), diagnostic of significant interdomain motion<sup>16</sup>. **c**, Depiction of interdomain motion in the FBP3/4–M29 complex. The red and blue cones indicate the slow motion of the KH4 and KH3 halves of the complex, respectively. The KH4 and KH3 domains are shown as red and blue ribbons, respectively, and the ssDNA bound to them is shown in purple and green, respectively. The 5-base linker for the ssDNA and 30-residue linker for the protein are depicted by blue and red dashed lines, respectively. The semi-cone angle  $\theta$ , derived from the internal slow order parameter  $S_s^2$  is about 30° for both domains.  $D_{||}$  and  $D_{\perp}$  represent the parallel and perpendicular components of an axially symmetric diffusion tensor. The two domains are aligned relative to the long axis of the diffusion tensor. The overall length of the complex and the separation between the two domains is indicated and calculated as described in the text. Nucleotide numbering is in italics.



determined by the intrinsic flexibility of the ssDNA (bases 11–15 of M29) linking the KH3 and KH4 binding sites. The 30-residue protein linker between the KH3 and KH4 domains is highly disordered (Fig. 3a), such that in free FBP3/4 the two domains reorient essentially independently of each other (unpublished data), as in the case of other modular proteins connected by long linkers<sup>16</sup>. Further, the ssDNA linker, although it is flexible, is not disordered because the observed NOEs are typical of right-handed DNA<sup>8</sup>.

The distances ( $L_i$ ) from the centre of KH3 and KH4 domains to the flexible hinge point at the centre of the overall complex can be derived from the relation  $L_i = (D_{\text{trans},i}/D_{\text{w},i})^{1/2}$ , where  $D_{\text{trans},i}$  and  $D_{\text{w},i}$  are the translational and rotational diffusion coefficients of domain  $i$  (ref. 19).  $D_{\text{trans},i}$  is calculated from the hydrodynamic radius ( $\sim 16.5$  Å for each domain) and  $D_{\text{w},i}$  from the values of  $S_0^2$  and  $\tau_s$  (ref. 19 and Supplementary Information). This yields  $D_{\text{trans},i} \approx 1.9 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ ,  $D_{\text{w},i} \approx 1.7 \times 10^7 \text{ s}^{-2}$ , and  $L_i \approx 33$  Å. The overall average length of the complex is therefore about 100 Å, corresponding to a cylinder length to diameter ratio of 3 with a predicted diffusion anisotropy of 1.8 (ref. 20), in agreement with the value determined from the model free analysis of the data. These dimensions imply that the average distance between the C terminus of KH3 (residue 74) and the N terminus of KH4 (residue 104) is about 35 Å, which is close to the expected average end-to-end distance of about 40 Å for a random-coil polypeptide of  $n = 30$  residues and a 40% Gly content<sup>20</sup>. The intervening stretch of ssDNA comprises six base steps, so the average rise for this segment of ssDNA is around 5.8 Å, about 70% longer than regular B-DNA.

Interdomain motion observed in the FBP3/4–M29 ssDNA complex in which the linkers, both protein and ssDNA, possess intrinsic flexibility, is likely to be an essential component of FBP function. This property can readily accommodate changes in the direction of the DNA. FBP interacts directly with and modulates the helicase activity of TFIIF<sup>4</sup>, a generalized transcription factor that has been proposed to act as a torque-generating machine<sup>21</sup> and is required throughout the early stages of transcription from initiation through promoter escape. Hence, interdomain flexibility in the FBP–ssDNA complex allows FBP bound to FUSE to maintain contact with a moving component, TFIIF, of the transcription machinery. As such, the ssDNA-binding domain of FBP presents an unusual architectural transcription factor: the presence of multiple KH domains, separated by linkers, generates a series of single-stranded flexible hinge points in the DNA matrix, facilitating interactions between other, sometimes mobile, components of the transcription machinery that would otherwise be prohibited by stereochemical and topological restraints imposed by the more rigid double-stranded DNA. This predicts that reduction in interdomain flexibility, by reducing the length of the intervening ssDNA sequence between the KH3- and KH4-binding sites, should decrease FUSE activity, whereas increasing the length of this linker segment slightly should have less effect as it could be accommodated by the flexible protein linker between the two KH domains. This is borne out experimentally: deletion of four of the five intervening bases between the KH4 and KH3 ssDNA binding sites (corresponding to bases 11–14 of M29) reduces *c-myc* reporter expression approximately fourfold<sup>22</sup>; conversely, insertion of an additional four bases between the two ssDNA binding sites has minimal effect on FUSE activity (D.L. and M. Avigan, unpublished work).

Underlying the structure of the FBP3/4–ssDNA complex is the concept that DNA transcription can be monitored and controlled by the recognition of ssDNA generated during the course of transcription. The present structure clearly shows that ssDNA located in a *cis*-regulatory element (FUSE) upstream from the *c-myc* promoter can be recognized specifically by FBP3/4, forming a high-affinity complex in solution. These features, combined with previous functional studies on the role of FBP at the *c-myc* locus<sup>1,4–6</sup>, provide the experimental basis for this mechanism of transcriptional control.

Finally, we note that overexpression of *c-myc* is associated with many human cancers<sup>23</sup> and that loss of FBP function shuts off *c-myc* expression and arrests cellular proliferation<sup>24</sup>. Thus, the FBP3/4–ssDNA complex may represent a potentially useful target for therapeutic intervention. □

## Methods

### Sample preparation

FBP3/4, comprising residues 278–447 of human FBP<sup>2</sup> and numbered from residues 5–174, as well as the individual KH3 (residues 5–77) and KH4 (residues 101–174) domains were cloned, expressed and purified by affinity chromatography using standard procedures. Samples for NMR contained 1:1 complexes of protein (<sup>15</sup>N, <sup>15</sup>N/<sup>13</sup>C or <sup>15</sup>N/<sup>13</sup>C/<sup>2</sup>H/VLI-methyl protonated) and ssDNA in 50 mM sodium phosphate, 20 mM EDTA and 0.02% NaN<sub>3</sub>, pH 6.8. Further details are provided in Supplementary Information.

### NMR spectroscopy

All NMR experiments were carried out at 35 °C on Bruker 600, 750 and 800 MHz spectrometers. <sup>1</sup>H, <sup>15</sup>H and <sup>13</sup>C backbone and side-chain resonances of the protein were assigned by three-dimensional (3D) double- and triple-resonance NMR experiments<sup>8</sup>; <sup>1</sup>H resonances for the ssDNA by two-dimensional (2D) <sup>13</sup>C-filtered experiments<sup>8</sup>; and <sup>31</sup>P resonances of the ssDNA from a 2D <sup>1</sup>H–<sup>31</sup>P correlation spectrum. The observation of strong <sup>31</sup>P(i)–H4'(i) correlations and the absence of <sup>31</sup>P(i)–H5'/H5''(i) correlations indicates that the β and γ sugar-phosphate torsion angles are in the *g* and *g*<sup>+</sup> conformations, respectively<sup>13</sup>. Interproton distance restraints (classified into ranges<sup>8</sup>) within the protein were derived from 3D and four-dimensional (4D) <sup>15</sup>N- and <sup>13</sup>C-separated nuclear Overhauser enhancement (NOE) experiments; within the DNA from 2D <sup>13</sup>C-filtered NOE experiments; and between the protein and DNA from 3D <sup>13</sup>C-separated/<sup>13</sup>C-filtered and <sup>15</sup>N-separated/<sup>13</sup>C-filtered NOE experiments<sup>8</sup>. Backbone φ/ψ torsion angle restraints were derived from backbone chemical shifts using the program TALOS<sup>25</sup>. Heteronuclear <sup>3</sup>J couplings were measured by quantitative J-correlation spectroscopy<sup>8</sup>. Dipolar couplings were measured in a dilute liquid crystalline medium of 25 mg ml<sup>−1</sup> phase fd<sup>15</sup>.

### <sup>15</sup>N relaxation measurements

<sup>15</sup>N T<sub>1</sub>, T<sub>2</sub> and NOE measurements were carried out at 600 and 750 MHz and analysed essentially as described<sup>18</sup>. All the relaxation data were fitted simultaneously on the basis of the atomic coordinates optimizing the value of the effective overall rotational correlation time and anisotropy for the whole complex; a single-order parameter  $S_0^2$  used to describe all fast internal motions; the order parameters ( $S_0^2$ ) and correlation times ( $\tau_s$ ) for slow interdomain motions of the two domains; and the two Euler angles describing the orientation of the long axis of the rotational diffusion tensor relative to the coordinates of each domain. The correlation time for fast internal motions,  $\tau_e$ , was held fixed at 15 ps, but the results are not affected for values of less than 30 ps (ref. 17). The optimized value of  $S_0^2$  is 0.85, typical of fast librational motions<sup>17–19</sup>. The value of the error function  $E/N$  is 3.3.

### Structure calculations

Structures were calculated from the experimental restraints by simulated annealing in torsion angle space<sup>26</sup> using XPLOR\_NIH ([http://nmr.cit.nih.gov/xplor\\_nih](http://nmr.cit.nih.gov/xplor_nih)). The non-bonded contacts in the target function are represented by a quartic van der Waals repulsion term supplemented by a torsion angle database potential of mean force<sup>27</sup>. Structure figures were generated with the programs VMD-XPLOR<sup>28</sup> and GRASP<sup>29</sup>.

Received 27 July; accepted 11 December 2001.

- Liu, J. *et al.* Defective interplay of activators and repressors with TFIIF in xeroderma pigmentosum. *Cell* **104**, 353–363 (2001).
- Rothman-Denes, L. B., Dai, X., Davydova, E., Carter, R. & Kazimierczak, K. Transcriptional regulation by DNA structural transitions and single-stranded DNA binding proteins. *Cold Spring Harbor Symp. Quant. Biol.* **63**, 63–73 (1999).
- Michelotti, E. F., Sanford, S. & Levens, D. Marking of active genes on mitotic chromosomes. *Nature* **388**, 895–899 (1997).
- Liu, J. *et al.* The FBP interacting repressor targets TFIIF to inhibit activated transcription. *Mol. Cell* **5**, 331–341 (2000).
- Duncan, R. C. *et al.* A sequence-specific, single strand binding protein activates the far-upstream of *c-myc* and defines a new DNA binding motif. *Genes Dev.* **8**, 465–480 (1994).
- Michelotti, G. A. *et al.* Multiple single-stranded *cis* elements are associated with activated chromatin of the human *c-myc* gene *in vivo*. *Mol. Cell. Biol.* **16**, 2656–2669 (1996).
- Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA binding proteins. *Science* **265**, 615–621 (1994).
- Clare, G. M. & Gronenborn, A. M. Determining structures of larger proteins and protein complexes by NMR. *Trends Biotechnol.* **16**, 22–34 (1998).
- Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332 (2000).
- Baber, J. L., Libutti, D., Levens, D. & Tjandra, N. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a *c-myc* transcription factor. *J. Mol. Biol.* **289**, 949–962 (1999).
- Musco, G. *et al.* Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* **85**, 237–245 (1996).
- Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).

13. Gorenstein, D. G. Conformation and dynamics of DNA and protein-DNA complexes by  $^{31}\text{P}$  NMR. *Chem. Rev.* **94**, 1315–1138 (1994).
14. Clore, G. M., Starich, M. R. & Gronenborn, A. M. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.* **120**, 10571–10572 (1998).
15. Clore, G. M., Gronenborn, A. M. & Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **131**, 159–162 (1998).
16. Braddock, D. T., Cai, M., Baber, J. L., Huang, Y. & Clore, G. M. Rapid identification of medium to large scale interdomain motion in modular proteins using dipolar couplings. *J. Am. Chem. Soc.* **123**, 8634–8635 (2001).
17. Clore, G. M. *et al.* Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112**, 4989–4990 (1990).
18. Baber, J. L., Szabo, A. & Tjandra, N. Analysis of slow interdomain motion of macromolecules using NMR relaxation data. *J. Am. Chem. Soc.* **123**, 3953–3959 (2001).
19. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear-magnetic-relaxation in macromolecules. 1 Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1980).
20. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry* Part II, 564 and Part III, 1008–1010 (Freeman, San Francisco, 1980).
21. Kim, T.-K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1421 (2000).
22. Avigan, M. L., Stober, B. & Levens, D. A far upstream element stimulates *c-myc* expression in undifferentiated leukemia cells. *J. Biol. Chem.* **265**, 18538–18545 (1990).
23. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16**, 653–699 (2000).
24. He, L. *et al.* Loss of FBP function arrests cellular proliferation and extinguishes *c-myc* expression. *EMBO J.* **19**, 1034–1044 (2000).
25. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database of protein chemical shift and sequence homology. *J. Biomol. NMR* **31**, 289–302 (1999).
26. Schwieters, C. D. & Clore, G. M. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J. Magn. Reson.* **152**, 288–302 (2001).
27. Clore, G. M. & Gronenborn, A. M. New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA* **95**, 5891–5898 (1998).
28. Schwieters, C. D. & Clore, G. M. The VMD-XPLOR visualization package for NMR structure refinement. *J. Magn. Reson.* **149**, 239–244 (2001).
29. Nicholls, A., Sharp, K. A. & Honig, G. Protein folding and association: insights into interfacial and thermodynamic properties of hydrocarbons. *Proteins* **17**, 297–309 (1991).
30. Clore, G. M. & Garrett, D. S. R-factor, free R and complete cross-validation for dipolar coupling refinement of NMR structures. *J. Am. Chem. Soc.* **121**, 9008–9012 (1999).

**Supplementary Information** accompanies the paper on *Nature's* website (<http://www.nature.com>).

## Acknowledgements

We thank D. Garrett and F. Delaglio for software support; C. A. Bewley, M. Caffrey, W. A. Eaton, J. Kuszewski, L. Murphy, C. Schwieters, A. Szabo and N. Tjandra for discussions; and C. A. Bewley for critically reading the manuscript. This work was supported in part by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.).

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.M.C. (e-mail: [clore@speck.niddk.nih.gov](mailto:clore@speck.niddk.nih.gov)). Coordinates have been deposited in the Protein Data Bank (RCSB accession code 1J4W).

## correction

# Transmission potential of smallpox in contemporary populations

Raymond Gani & Steve Leach

*Nature* **414**, 748–751 (2001).

There was an error in equations (1) in this Letter. The  $\theta I$  terms in  $dI/dt$  and  $dQ/dt$  should have been omitted. Thus, the set of differential equations should have been:

$$\begin{aligned} \frac{dS}{dt} &= \chi_1(1 - \epsilon_1)C_i - \beta(\varphi + \rho - \varphi\rho)SI & \frac{dI}{dt} &= \alpha(1 - \theta)E_n - \gamma I \\ \frac{dE_n}{dt} &= \beta\varphi(1 - \rho)SI - \alpha E_n & \frac{dQ}{dt} &= \alpha(1 - \epsilon_2)E_i + \alpha\theta E_n - \chi_2 Q \\ \frac{dE_i}{dt} &= \beta\varphi\rho SI - (\chi_1\epsilon_2 + \alpha(1 - \epsilon_2))E_i & \frac{dU}{dt} &= \gamma I + \chi_2 Q \\ \frac{dC_i}{dt} &= \beta\rho(1 - \varphi)SI - \chi_1 C_i & \frac{dV}{dt} &= \chi_1(\epsilon_2 E_i + \epsilon_1 C_i) \end{aligned} \quad (1)$$

and the definition of  $\theta$  in Table 2 should have read ‘the proportion of infectious individuals from  $E_n$  that are quarantined’. This change only affects the analysis of the data for Kosovo. The value of  $R_0$  that gives the best fit to the data for Kosovo is now 10 (5 discounting hospital-associated cases), rather than 10.8 (5.4 discounting hospital-associated cases), as stated in the original paper in Table 1. This small error does not alter our conclusions.  $\square$

# Getting into the swing of things

Focusing on centrifuges and general lab equipment.

## PepMap

**Dionex**

[www.dionex.com](http://www.dionex.com)

Columns for TFA-free LC/MS

PepMap is a silica-based stationary phase in which virtually all silanophobic activities are suppressed. It is available on columns with an I.D. ranging from 50  $\mu$ m to 1.0 mm. Use of the PepMap material facilitates the separation of peptides with minimal band-broadening. In addition, a 10-fold lower concentration of TFA can be used in the mobile phase; TFA concentration generally needs to be only 0.005% to achieve symmetric peaks. It is also possible to exchange the TFA with acetic acid 0.05%. In both cases, Dionex claims, an increased signal-to-noise ratio is achieved. PepMap is suitable for separation of peptide mixtures obtained by, for example, tryptic digestion of proteins, particularly when using a mass spectrometer as detector.

## A-4-81 centrifuge rotor

**Eppendorf**

[www.eppendorf.com](http://www.eppendorf.com)

In full swing

Eppendorf's A-4-81 swing-bucket rotor increases the load capacity of the company's 5810 and 5810 R centrifuges to 1,600 ml. The rectangular buckets can be used to centrifuge at a maximum speed of 4,000 litres per minute (maximum RCF 3,260g). The rotor is compatible with most commonly available tubes with volumes of between 1.5 and 400 ml, including Vacutainer or glass test-

tubes,



Load carrier: the A-4-81 swing-bucket rotor.

and up to twenty 50-ml Falcon tubes. Accessories include a modular adapter system and covers for aerosol-tight centrifugation. All rotors, buckets, covers and adapters are autoclavable.

## 2028R multipurpose centrifuge

**Napco**

[www.napco2.com](http://www.napco2.com)

Quick change

The 2028R multipurpose centrifuge from Napco comes with a selection of swing-out and angle rotors and is capable of both high- and low-speed centrifugation. Rotors glide into position on the drive shaft and lock automatically, without the use of traditional nuts and wrenches. Napco says this feature allows rotor exchanges to be completed in less than five seconds.

## Heraeus HERAfreeze

**Kendro**

[www.kendro.com](http://www.kendro.com)

Ten out of ten for sample storage

Kendro offers six chest models and four low-profile upright models in its range of ultra-low-temperature freezers, providing storage capacities ranging from 3.0 to 24.4 cubic feet. Top Series models feature microprocessor control with digital set point and display, alarm and monitoring capabilities and enhanced control functions. Basic Series freezers incorporate a solid-state control

system with digital set point display and pre-set audible alarm. The cabinet insulation is a



Versatility from Napco.

minimum depth of 5 inches and the security latch system includes a sturdy handle, keyed lock and provision for a padlock. Doors can be opened and closed with one hand. The freezers can be used with a wide range of standard or custom inventory racks, boxes, dividers, canisters and other storage systems.

## PlateManager and GelManager

**Zumatrix**

[www.zumatrix.com](http://www.zumatrix.com)

Make tracks for molecular biology

These new software modules will facilitate automation of testing in DNA laboratories. They are fully integrated with the company's Matrix Plus LIMS and provide support for sample processing from registration, receipt, extraction, purification, amplification, identification and reporting. The contents of plate wells and gel plates are recorded and tracked throughout the entire laboratory process. Status and audit reports are available at any time.

## IN20 chilling/heating incubator

**Torrey Pines Scientific**

[www.torrey-pinescientific.com](http://www.torrey-pinescientific.com)

A coloured view of sample incubation

Measuring 8 inches wide by 14.5 inches deep, the IN20 Peltier-based chilling/heating incubator from Torrey Pines Scientific is made of emerald green translucent acrylic for easy sample viewing, and comes with six shelves, each holding two assay plates. Two of the shelves have 192 holes for 0.2-ml centrifuge or PCR tubes in 8- or 12-tube strips; a further two shelves each have 80 holes for 0.5-ml centrifuge tubes; and the remaining two shelves come with 60 holes for 1.5-ml centrifuge tubes. The IN20 has no compressors or CFCs. It has a two-line, backlit alphanumeric LCD that reads





**The IN-20 incubator: in-car entertainment?**

temperature to 0.1°C and a 30-day timer. A built-in data logger and RS232C interface provide control or data logging onto a PC. The incubator can run on 12 volts d.c. from the desktop power supply provided, or, with an accessory cable, can be run from a car cigarette lighter for field operation or sample transport. Applications include culture growth above, below or at ambient temperature, protein crystal growth, storing oocytes and DNA libraries and hybridizations.

### Purifier horizontal clean bench

**Labconco** [www.labconco.com](http://www.labconco.com)  
*Clear the air*

The Purifier workstation provides a particulate-free Class 100 environment by drawing room air through a prefilter in the top of the cabinet to trap large particles; the air then flows through a 99.99% efficient HEPA filter and is projected horizontally across the entire work area, preventing cross-contamination. The bench features a double gasket and negative pressure plenum that surround the HEPA filter to provide primary and secondary leak protection. The motor/blower is installed on vibration-dampening mounts. For extremely vibration-sensitive tasks such as those involving microscopes, a vibration isolation table is available. The bench is available in 3-, 4- and 6-inch widths. Applications include plant tissue culture research, media preparation, electronic part inspection, syringe filling and medical device assembly.

### AdenoVator, AdenoExpress, GeneSHUTTLE

**Qbiogene** [www.qbiogene.com](http://www.qbiogene.com)  
*Adenovirus solutions*

Qbiogene has launched three new adenovirus products. The AdenoVator adenoviral vector system enables the user to construct adenoviruses in *E. coli*, obtain a high level of protein expression and co-express a gene of interest with GFP or BFP. AdenoExpress ready-made,

pre-qualified, pre-titrated adenoviruses are available for expressing a variety of human and mouse genes. GeneSHUTTLE-20 reagent is an all-purpose transfection reagent for primary cells such as PC-12 where efficiencies of 45–50% can be obtained.

### GeneAmp

**Applied Biosystems**  
[www.appliedbiosystems.com](http://www.appliedbiosystems.com)

### High-fidelity PCR

GeneAmp can amplify up to 10 kb of DNA and, according to Applied Biosystems, is ideal for applications such as cloning and site-directed mutagenesis. AmpliTaq DNA polymerase is blended with a new thermostable proofreading enzyme to increase PCR fidelity threefold and improve the yield of Taq DNA polymerase. The system comes with a high-fidelity PCR buffer to enhance enzyme activity when used in combination with the enzyme mix. GeneAmp is designed for use with the GeneAmp PCR system 9700, 9600, 2700 or 2400 and can be used with a wide range of template samples, including mammalian genomic DNA, artificial chromosomes, plasmids and post-reverse transcription.

### Swinging bucket rotors

**Beckman Coulter** [www.beckmancoulter.com](http://www.beckmancoulter.com)  
*For DNA prep*

The new JS-5.3 four-place rotor for Beckman Coulter's Avanti J-20XP centrifuge processes up to 24 microplates at 6,130g, while the JS-5.9 two-place rotor handles up to ten 96-well plates in either the Avanti J-25 or Avanti

J-301 centrifuges at up to 6,530g. The rotors are suitable for plasmid purification, dye terminator removal, ethanol precipitation, cell culture pelleting, cell fractionation and protein precipitation. Accessories include 96-well plates, deep-well plates, square-well plates and DNA kit 96-well filtration plates. High-throughput processing kits are available for DNA chemistries such as Qiagen, Amicon and Promega.

### ImmobaSil

**Ashby Scientific** [www.ashby-scientific.co.uk](http://www.ashby-scientific.co.uk)  
*Cell support matrix*

ImmobaSil supports the growth of both adherent and suspension cells in culture. It has a porous structure that provides a large surface area. The gas-permeable silicon matrix enhances oxygen transfer, extending the life of long-term cultures. Even when cells have fully colonized the carrier, the design of the pores and macro-geometry ensure optimal nutrient flow, allowing high cell densities to be achieved. Inert and non-toxic, ImobaSil is biocompatible with even the most delicate mammalian cell lines. Cell attachment times are typically two hours with cell densities in excess of  $1 \times 10^8$  per ml of carrier. The matrix is supplied pre-wetted and is available in a range of formats suitable for use in a wide variety of fermenters and culture vessels, including spinner flasks, stirrer tanks and fluidized or packed bed reactors. It can be repeatedly sterilized with steam or chemical reagents.

*These notes are compiled in the Nature office from information provided by the manufacturers.*



**Beckman Coulter's two- and four-place rotors, aimed at DNA preparation.**

## Contacts

**Publisher:** Fabien Savenay  
**Editor:** Paul Smaglik  
**Sales Director:** Ben Crowe

## European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

## Group European Manager:

Leonie Welss (4954)  
**European Manager:**  
Nevin Bayoumi (4978)

## UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),  
Andy Douglas (4975)  
**Holland/ Italy:** Nevin Bayoumi (4978)  
**Scandinavia:** Silke Opstrup (4994)  
**Spain/ Portugal:** Leonie Welss  
(4954)

## Production Manager:

Billie Franklin  
To send films and materials use  
London address above.  
Tel +44 (0) 20 7843 4814  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

## International

## Advertising Coordinator:

Laura Pearson (4977)  
**Naturejobs web development:**  
Tom Hancock  
**Naturejobs online production:**  
Ben Lund

## European Satellite Offices

## France/ Belgium:

Christine Niox-Chateau  
Tel + 33 (0) 1 43 20 16 51  
Fax + 33 (0) 1 43 20 51 52  
e-mail: [c.nioxchateau@nature.com](mailto:c.nioxchateau@nature.com)

## Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner  
Tel + 49 89 54 90 57 11/-2  
Fax + 49 89 54 90 57 20  
e-mails: [p.phelan@nature.com](mailto:p.phelan@nature.com)  
[k.turner@nature.com](mailto:k.turner@nature.com)

## US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707  
Tel +1 800 989 7718  
Fax +1 800 989 7103  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

## US Sales Director:

Ben Crowe  
**US Sales Manager:** Peyton Mason

## US Advertising Coordinator:

Ashly de Leon

## Japan Head Office, Tokyo

MG Ichigaya Building (5F),  
19-1 Harakata-machi,  
Shinjuku-ku,  
Tokyo 162-0841  
Tel +81 3 3267 8751  
Fax +81 3 3267 8746  
e-mail: [k.cowan@naturejp.com](mailto:k.cowan@naturejp.com)  
**Japan Manager:** Kate Cowan

# naturejobs

## Money worries

Last week, thousands of students marched through London campaigning for the abolition of tuition fees and the restoration of full government-funded grants. The protest followed complaints by the National Union of Students that many undergraduates are living on or near the poverty line. Finding money to pay both rent and tuition fees forces many students to take out loans — and leaves them with the opinion that they would be better off on the dole.

A spokesperson at the UK Department for Education and Skills countered these claims as untrue: “Graduates can expect to earn on average 35% more than the average — that’s an extra £400,000 (US\$570,000) over a lifetime.”

For scientific employment, reality falls somewhere in between the competing claims. Yes, graduates do eventually get higher salaries, but being heavily in debt at the end of a degree can steer people away from some career choices.

There are already some hints that this is happening. David James, chair of the department of pharmacology at the University of Cambridge, believes there are signs that students are shying away from academic careers. One of his postdocs, who was among the first wave of English undergraduates required to pay tuition fees, said that he could not consider seeking a professorship because he needed a higher income to pay off his loans. Instead, he decided to go into industry (see *Naturejobs* 4–5; 1 November 2001).

With fewer students enrolling in maths, physics and chemistry — and a wave of academic retirements on the horizon — it would seem wise to encourage students to pursue scientific subjects, which means making academic careers look as though they might be a feasible option.

## Paul Smaglik

*Naturejobs* editor



## Contents

### POSTDOCS & STUDENTS

Teaching postdocs to teach p5

### MOVERS

Joint Genome Institute director heads off for pastures new; Collège de France gets second chemistry chair; QB3 gains as NIH loses; and more Back page

### WWW.NATUREJOBS.COM

Career centre  
Information on the scientific job market

## FOCUS

## SPOTLIGHT

## RECRUITMENT

## SCIENTIFIC ANNOUNCEMENTS

## SCIENTIFIC EVENTS

# At last, a chance for postdocs to learn how to teach

Gordon Uno:

The struggle is whether you have the 'luxury' of having a teaching postdoc or someone who might be focusing more on research.



**M**ost postdocs have little teaching experience when they apply for their first faculty position. They have been primed to emphasize research, often at the expense of everything else. So when they enter a tenure-track faculty position that includes teaching responsibilities, they are often unprepared to develop and teach a curriculum.

"They're caught," says Gordon Uno, chairman of the department of botany and microbiology at the University of Oklahoma. The educational system has not been structured to give postdocs the teacher training they need. But a few postdoctoral programmes are now being designed to emphasize teaching. Some are funded by universities, others are sponsored by the US National Institutes of Health (NIH), and others have been created by foundations.

In 1999, Uno's department created and part-funded two positions for 'teaching postdocs'. These were to teach classes, improve introductory botany and microbiology classes, and bring more technology into the classroom. Usually, though, this type of postdoc just crops up department by department, as needed.

"I don't see a trend sweeping the country," says Uno. "The struggle is whether you have the 'luxury' of having a teaching postdoc as opposed to someone who might be focusing more on research."

Teaching postdocs may attract scientists who want to work at institutions that only grant a first degree. But they might find some lingering arrogance about their pedigree if they applied for positions at PhD-granting universities.

## INDEPENDENT ANSWER

Although most teaching postdocs are scattered throughout university departments in the United States, a few institutions and foundations are taking a more organized approach. In 1999, the National Institute for General Medical Sciences (NIGMS) launched a five-year, \$3.1-million programme aimed at both educating postdocs for academic positions and increasing minority representation in the sciences.

Under the programme, called the Institutional Research and Academic Career Development Award, postdocs at four universities — Emory, North Carolina,

Arizona and the University of California, Davis — spend part of their time teaching at institutions whose students are mainly members of ethnic minorities. Salaries, starting at \$28,260, are based on NIH standards (see *Naturejobs* 5; 10 January 2002).

Each grant emphasizes different ways to gain teaching experience. For example, the Davis–San Francisco State University (SFSU) programme, Professors of the Future (PROF), requires research collaboration between faculty mentors at SFSU and Davis. The postdocs are coached in teaching, they co-teach with their SFSU mentor and finally take a course alone. "Teaching is part of what they're here for and it's not an extracurricular activity," says Jerry Hedrick, professor of biochemistry at Davis and programme director for the Davis–SFSU partnership.

## CHEMISTRY ESTABLISHED

Another established programme is the Scholar/Fellow Program for Undergraduate Institutions, set up in 1987 and supported by the Camille and Henry Dreyfus Foundation. So far, 200 fellows have completed the programme of teaching and research in chemistry, chemical engineering and biochemistry departments at undergraduate institutions. Salaries start at a competitive \$35,000.

Patricia Reggio, professor of chemistry and biochemistry at Kennesaw State University, Georgia, has just received her first Scholar/Fellow award. She hired a postdoc to work on computational modelling of drug-receptor interactions, teach a lecture course in general chemistry and develop a senior-level course in computational chemistry. "The idea is for the faculty-award recipient to mentor the postdoc," says Reggio.

Brian Coppola, an associate professor of chemistry at the University of Michigan in Ann Arbor, has also sponsored several teaching postdocs in the past. Now the University of Michigan has started a teaching postdoc programme similar to the Davis–SFSU arrangement with Oberlin and Kalamazoo colleges.

Postdocs who can teach are a move in the right direction for future faculty, says Coppola, "so they don't walk into the first day of their professional career as unprepared as has been historically true". ■

Karen Kreeger is a freelance science writer based in Philadelphia.

## Web links

Camille and Henry Dreyfus Foundation  
 ♦ [www.dreyfus.org](http://www.dreyfus.org)  
 National Institute for General Medical Sciences  
 ♦ [www.nigms.nih.gov](http://www.nigms.nih.gov)  
 PROF  
 ♦ [gradstudies.ucdavis.edu/postdocs/jhpost.htm](http://gradstudies.ucdavis.edu/postdocs/jhpost.htm)



## Transitions

San Diego-based structural proteomics company Syrrx last month announced the appointment of **Keith Wilson** as vice-president of technology. Wilson was previously a project leader with Vertex Pharmaceuticals of Cambridge, Massachusetts.

Former NASA astronaut and assistant deputy administrator **Charles Bolden** has been nominated to become the next deputy administrator for the space agency. Major General Bolden currently serves as the commanding general for the Third Marine Aircraft Wing.

Exelixis this month appointed **Robert Myers** as executive vice-president, pharmaceuticals. Myers comes to the South San Francisco-based biotech firm from ALZA, a drug-discovery company based in Mountain View, California. ALZA recently merged with Johnson & Johnson.

**Peter Freeman**, dean of computing at Georgia Institute of Technology, will join the US National Science Foundation in May as assistant director for computer and information science and engineering.

**Philip Davies**, professor of American studies at De Montfort University and chair of the British Association for American Studies, has been appointed the next director of the David and Mary Eccles Centre for American Studies at the British Library.

**Frank Bash**, director of the McDonald Observatory at the University of Texas at Austin, has announced that he will leave his post on 31 August 2003. The search for his replacement will begin immediately.

## BIOTECHNOLOGY

After almost 27 years with the US National Institute of General Medical Sciences, director **Marvin Cassman** is leaving. In May, he will become head of the Institute for Bioengineering, Biotechnology and Quantitative Biomedical Research (QB3). Currently under construction, QB3 will be based at the University of California, San Francisco, and will be shared with the university's Berkeley and Santa Cruz campuses.

Cassman is leaving the US National Institutes of Health in the wake of several other high-level departures, but says that the timing is coincidental. "I'm not going because I'm tired or unhappy here," he says. Instead, Cassman says the move will allow him to focus more strongly on one of his major interests: integrating physics, mathematics and biology into a systems approach to biology.

He anticipates that the "key problem" in building the new institute will be coordinating research activities across the three separate campuses. He notes that arranging communication between the three, which are spread throughout the Bay Area, is just as problematic as coordinating sites dispersed around the country, as commuting from campus to campus can take hours because of the area's notorious traffic problems.

## CHEMISTRY

Last month, **Jacques Livage** delivered his first lecture for his chair in condensed matter chemistry at the Collège de France. It marked the first time since 1930 that two chemists have held such positions at the Collège de France. The other chemistry chair is held by **Jean-Marie Lehn**, who shared the chemistry Nobel prize in 1987.

Livage has been a professor of chemistry at the Pierre and Marie Curie University in Paris since 1973, but his new position means that he will no longer handle administrative duties there. Instead, he will continue with his lab research, and will prepare and deliver 18 lectures a year around the country. "We're not allowed to give the same lecture twice," Livage says. That challenge will allow him to present up-to-date scientific material to his audiences. "It's a very pleasant thing to give lectures on an interesting topic," he says, "and you have no exams."

The timeliness of his lecture subjects will help them to meld with his research interests in solid-state chemistry — especially *chimie douce* ('soft chemistry'), which he helped to pioneer with **Jean Rouxel** in the 1970s. Livage is keen on



Marvin Cassman



Jeffrey Nye



Jacques Livage

making hybrid materials by using soft-solution processing approaches to mix organic and inorganic starting materials. His lab is also working on trapping bacteria inside silica gels — an approach that he hopes will result in new types of biosensor and drug-delivery systems.

## PHARMACEUTICALS

**Jeffrey Nye** paved the way for his transition from academia to industry through a steady stream of consulting work. Nye left the Northwestern University Medical School in Chicago late last year to take up a post with drug company Pharmacia in Kalamazoo, Michigan. But he has been advising drug-development companies and venture capitalists since he was a graduate student. While at Northwestern, Nye's work focused on the genetics of spina bifida.

As director of central nervous system-related genomics discovery for Pharmacia, Nye is enthusiastic about the potential of therapeutics for Alzheimer's disease. The chance to translate basic research on that disease into potential treatment was one attraction that lured him to the job. He also enjoys being involved in business, medicine, science and law. "Here I get to merge all my interests," he says of his new post.

## GENOMICS

Life-sciences company Amersham Biosciences this month announced the appointment of **Trevor Hawkins** as senior vice-president of genomics. Hawkins joins Amersham from his position as director of the US Department of Energy's Joint Genome Institute (JGI).

When he joined the JGI as deputy director in 1999, its funding and mission were unclear. But that all changed when the JGI played an important role in sequencing the human genome — an accomplishment that Hawkins found gratifying. "It's always nice to leave on a high," he says.

In some respects, his departure from the institute is a return to his commercial roots. Before joining the JGI, Hawkins was vice-president of genomics for CuraGen, a biotech company in New Haven, Connecticut. There he had an almost entirely scientific role. At Amersham, he is looking to branch out, taking on business development and marketing responsibilities. "I had always seen my future as running a commercial enterprise," he says. "This is a good step forward for my own career goals."

## CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:  
 ◆ [naturejobseditor@naturedc.com](mailto:naturejobseditor@naturedc.com)